

Input file Fbh8099FL.tra; Output File Fbh8099FL.tra
Sequence length 2725

CCACGCGTCCGGCCTTCCGAAATAGAAACAAAGTTGGTCACAAATCACATTAGCTTTGCCCGAAGTTTTTCCCCACACT

CTTCTTTAGCATGCTATTATGGGGAAAGTGACCACTCCTGGGAGCGGGGTGGTCGGGGCGGTTTGGTGGCGGGGAAGC

	M	V	P	V	E	N	T	E	G	P	S	L	L	N	14			
GGCTGTA	ACTTCT	TACGT	GACC	ATG	GTA	CCT	GTT	GAA	AAC	ACC	GAG	GGC	CCC	AGT	CTG	CTG	AAC	42

Q	K	G	T	A	V	E	T	E	G	S	G	S	R	H	P	P	W	A	R	34
CAG	AAG	GGG	ACA	GCC	GTG	GAG	ACG	GAG	GGC	AGC	GGC	AGC	CGG	CAT	CCT	CCC	TGG	GCG	AGA	102

G	C	G	M	F	T	F	L	S	S	V	T	A	A	V	S	G	L	L	V	54
GGC	TGC	GGC	ATG	TTT	ACC	TTC	CTG	TCA	TCT	GTC	ACT	GCT	GCT	GTC	AGT	GGC	CTC	CTG	GTG	162

G	Y	E	L	G	I	I	S	G	A	L	L	Q	I	K	T	L	L	A	L	74
GGT	TAT	GAA	CTT	GGG	ATC	ATC	TCT	GGG	GCT	CTT	CTT	CAG	ATC	AAA	ACC	TTA	TTA	GCC	CTG	222

S	C	H	E	Q	E	M	V	V	S	S	L	V	I	G	A	L	L	A	S	94
AGC	TGC	CAT	GAG	CAG	GAA	ATG	GTT	GTG	AGC	TCC	CTC	GTC	ATT	GGA	GCC	CTC	CTT	GCC	TCA	282

L	T	G	G	V	L	I	D	R	Y	G	R	R	T	A	I	I	L	S	S	114
CTC	ACC	GGA	GGG	GTC	CTG	ATA	GAC	AGA	TAT	GGA	AGA	AGG	ACA	GCA	ATC	ATC	TTG	TCA	TCC	342

C	L	L	G	L	G	S	L	V	L	I	L	S	L	S	Y	T	V	L	I	134
TGC	CTG	CTT	GGA	CTC	GGA	AGC	TTA	GTC	TTG	ATC	CTC	AGT	TTA	TCC	TAC	ACG	GTT	CTT	ATA	402

V	G	R	I	A	I	G	V	S	I	S	L	S	S	I	A	T	C	V	Y	154
GTG	GGA	CGC	ATT	GCC	ATA	GGG	GTC	TCC	ATC	TCC	CTC	TCT	TCC	ATT	GCC	ACT	TGT	GTT	TAC	462

I	A	E	I	A	P	Q	H	R	R	G	L	L	V	S	L	N	E	L	M	174
ATC	GCA	GAG	ATT	GCT	CCT	CAA	CAC	AGA	AGA	GGC	CTT	CTT	GTG	TCA	CTG	AAT	GAG	CTG	ATG	522

I	V	I	G	I	L	S	A	Y	I	S	N	Y	A	F	A	N	V	F	H	194
ATT	GTC	ATC	GGC	ATT	CTT	TCT	GCC	TAT	ATT	TCA	AAT	TAC	GCA	TTT	GCC	AAT	GTT	TTC	CAT	582

G	W	K	Y	M	F	G	L	V	I	P	L	G	V	L	Q	A	I	A	M	214
GGC	TGG	AAG	TAC	ATG	TTT	GGT	CTT	GTG	ATT	CCC	TTG	GGA	GTT	TTG	CAA	GCA	ATT	GCA	ATG	642

Y	F	L	P	P	S	P	R	F	L	V	M	K	G	Q	E	G	A	A	S	234
TAT	TTT	CTT	CCT	CCA	AGC	CCT	CGG	TTT	CTG	GTG	ATG	AAA	GGA	CAA	GAG	GGA	GCT	GCT	AGC	702

K	V	L	G	R	L	R	A	L	S	D	T	T	E	E	L	T	V	I	K	254
AAG	GTT	CTT	GGA	AGG	TTA	AGA	GCA	CTC	TCA	GAT	ACA	ACT	GAG	GAA	CTC	ACT	GTG	ATC	AAA	762

S	S	L	K	D	E	Y	Q	Y	S	F	W	D	L	F	R	S	K	D	N	274
TCC	TCC	CTG	AAA	GAT	GAA	TAT	CAG	TAC	AGT	TTT	TGG	GAT	CTG	TTT	CGT	TCA	AAA	GAC	AAC	822

M	R	T	R	I	M	I	G	L	T	L	V	F	F	V	Q	I	T	G	Q	294
ATG	CGG	ACC	CGA	ATA	ATG	ATA	GGA	CTA	ACA	CTA	GTA	TTT	TTT	GTA	CAA	ATC	ACT	GGC	CAA	882

P	N	I	L	F	Y	A	S	T	V	L	K	S	V	G	F	Q	S	N	E	314
CCA	AAC	ATA	TTG	TTC	TAT	GCA	TCA	ACT	GTT	TTG	AAG	TCA	GTT	GGA	TTT	CAA	AGC	AAT	GAG	942

A	A	S	L	A	S	T	G	V	G	V	V	K	V	I	S	T	I	P	A	334
GCA	GCT	AGC	CTC	GCC	TCC	ACT	GGG	GTT	GGA	GTC	GTC	AAG	GTC	ATT	AGC	ACC	ATC	CCT	GCC	1002

Figure 1A

T	L	L	V	D	V	G	S	K	T	F	L	C	I	G	S	S	V	M	354	
ACT	CTT	CTT	GTA	GAC	CAT	GTC	GGC	AGC	AAA	ACA	TTC	CTC	TG	TT	GGC	TCC	TCT	GTG	ATG	1062
A	A	S	L	V	T	M	G	I	V	N	L	N	I	H	M	N	F	T	H	374
GCA	GCT	TCG	TTG	GTG	ACC	ATG	GGC	ATC	GTA	AAT	CTC	AAC	ATC	CAC	ATG	AAC	TTC	ACC	CAT	1122
I	C	R	S	H	N	S	I	N	Q	S	L	D	E	S	V	I	Y	G	P	394
ATC	TGC	AGA	AGC	CAC	AAT	TCT	ATC	AAC	CAG	TCC	TTG	GAT	GAG	TCT	GTG	ATT	TAT	GGA	CCA	1182
G	N	L	S	T	N	N	N	T	L	R	D	H	F	K	G	I	S	S	H	414
GGA	AAC	CTG	TCA	ACC	AAC	AAC	AAT	ACT	CTC	AGA	GAC	CAC	TTC	AAA	GGG	ATT	TCT	TCC	CAT	1242
S	R	S	S	L	M	P	L	R	N	D	V	D	K	R	G	E	T	T	S	434
AGC	AGA	AGC	TCA	CTC	ATG	CCC	CTG	AGA	AAT	GAT	GTG	GAT	AAG	AGA	GGG	GAG	ACG	ACC	TCA	1302
A	S	L	L	N	A	G	L	S	H	T	E	Y	Q	I	V	T	D	P	G	454
GCA	TCC	TTG	CTA	AAT	GCT	GGA	TTA	AGC	CAC	ACT	GAA	TAC	CAG	ATA	GTC	ACA	GAC	CCT	GGG	1362
D	V	P	A	F	L	K	W	L	S	L	A	S	L	L	V	Y	V	A	A	474
GAC	GTC	CCA	GCT	TTT	TTG	AAA	TGG	CTG	TCC	TTA	GCC	AGC	TTG	CTT	GTT	TAT	GTT	GCT	GCT	1422
F	S	I	G	L	G	P	M	P	W	L	V	L	S	E	I	F	P	G	G	494
TTT	TCA	ATT	GGT	CTA	GGA	CCA	ATG	CCC	TGG	CTG	GTG	CTC	AGC	GAG	ATC	TTT	CCT	GGT	GGG	1482
I	R	G	R	A	M	A	L	T	S	S	M	N	W	G	I	N	L	L	I	514
ATC	AGA	GGA	CGA	GCC	ATG	GCT	TTA	ACT	TCT	AGC	ATG	AAC	TGG	GGC	ATC	AAT	CTC	CTC	ATC	1542
S	L	T	F	L	T	V	T	D	L	I	G	L	P	W	V	C	F	I	Y	534
TCG	CTG	ACA	TTT	TTG	ACT	GTA	ACT	GAT	CTT	ATT	GGC	CTG	CCA	TGG	GTG	TGC	TTT	ATA	TAT	1602
T	I	M	S	L	A	S	L	L	F	V	V	M	F	I	P	E	T	K	G	554
ACA	ATC	ATG	AGT	CTA	GCA	TCC	CTG	CTT	TTT	GTT	GTT	ATG	TTT	ATA	CCT	GAG	ACA	AAG	GGA	1662
C	S	L	E	Q	I	S	M	E	L	A	K	V	N	Y	V	K	N	N	I	574
TGC	TCT	TTG	GAA	CAA	ATA	TCA	ATG	GAG	CTA	GCA	AAA	GTG	AAC	TAT	GTG	AAA	AAC	AAC	ATT	1722
C	F	M	S	H	H	Q	E	E	L	V	P	K	Q	P	Q	K	R	K	P	594
TGT	TTT	ATG	AGT	CAT	CAC	CAA	GAA	GAA	TTA	GTG	CCA	AAA	CAG	CCT	CAA	AAA	AGA	AAA	CCC	1782
Q	E	Q	L	L	E	C	N	K	L	C	G	R	G	Q	S	R	Q	L	S	614
CAG	GAG	CAG	CTC	TTG	GAG	TGT	AAC	AAG	CTG	TGT	GGT	AGG	GGC	CAA	TCC	AGG	CAG	CTT	TCT	1842
P	E	T	*																	618
CCA	GAG	ACC	TAA																	1854

TGGCCTCAACACCTTCTGAACGTGGATAGTGCCAGAACACTTAGGAGGGTGTCTTTGGACCAATGCATAGTTGCGACTC
 CTGTGCTCTCTTTTCAGTGTCTATGGAACCTGGTTTTTGAAGAGACACTCTGAAATGATAAAGACAGCCTTTAATCCCCCTC
 CTCCCCAGAAGGAACCTCAAAAGGTAGATGAGGTACAAGGTCTTAAGTGATCTCTTTTTCTGAGCAGGATATCAGGTTA
 AAAAAAAAAAGTTACTGGCTGGTTTAATACTTTCTACCTTCTTCACAGAGCAGCCTTTGAATAGACTATGTCCTAGTGA
 AGACATCAACCTCCGCCTTAAGCTATGTATGTATGGAGGCCAGTCGCAGCTTTATTATGCAGACACACAAGTGGTCTGG
 ACATGAGGGTACAGTTTCTGCCTACCAAGACACTACTTGCACCTGGATCTTACGCAAAAAAGAACCAGAACACACAGTGT
 GGACAACCTGCCCATATATTTCTATCTAGATTAGGAGAGGGTCTGGCTAGGATTTTAGTGGTAATTCCTAGTTACATTCA
 ACAAGTATAAAGATTATAGAGCTTATTTTATGAACATAAACTATAATTTAATGCAAAATATCCTTTTATGAATTTTCAT
 GTTAATATTGTGAAATATTAAAAATAATTCGCAATAAAAAAAAAAAAAAAAAAGGGCGGCCCGC

Figure 1B

Protein Family / Domain Matches, HMMer version 2

Searching for complete domains in PFAM

hmmpfam - search a single seq against HMM database

HMMER 2.1.1 (Dec 1998)

Copyright (C) 1992-1998 Washington University School of Medicine

HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /prod/ddm/seqanal/PFAM/pfam6.4/Pfam

Sequence file: /prod/ddm/wspace/orfanal/oa-script.8089.seq

Query: 8099

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
sugar_tr 1	Sugar (and other) transporter	318.2	9.6e-92	
FecCD_family 1	FecCD transport family	-218.2	6.9	
MCT 1	Monocarboxylate transporter	-235.8	2.7	

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
FecCD_family	1/1	26	227 ..	1	311 []	-218.2	6.9
sugar_tr	1/1	43	564 ..	1	488 []	318.2	9.6e-92
MCT	1/1	29	567 ..	1	611 []	-235.8	2.7

Alignments of top-scoring domains:

FecCD_family: domain 1 of 1, from 26 to 227: score -218.2, E = 6.9

```

*->GalsispadlvqalfgggtegeievdliiwdltlrRLPRvLlAlLV
      G+++ ++a  ++  + ++      ++      + +LV
8099  26  GSRHPPWARGCGMFT---FLSS-----VTAA-----VSGLLV 54

      GAaLAVaGAILQgltrNPLAsPgilGinsGAslgvvlaiivlfpgglisaisa
      G      + lGi sGA l +  ++l +  + ++
8099  55  G-----YELGIISGALLQIKTLLALSCH-EQEMV 82

      lylpsfAfaGaliallVyllawkgrnglspvrLiLaGialsalfsAlt
      +l+++A++ +l+++l+  +++      + i+ls+++ +l
8099  83  VSSLVIGALLASLTGGVLIDRYGRR-----TAILSSCLLGLG 120

      tlllllsddldqgqalfWltGSlsgrnWedvklalpilliglplalllar
      +l l+ls +  +++      + gr  v + l  ++ +  +a +
8099  121 SLVLILSLSYTVL-----IVGRIAIGVSISLSSIATCVYIAEI--- 158

      qLnvLsLGddtAkgLGvnvervR.llllllvvlLtGaaVAVAGpIgFVGL
      +++ R+ll++l  +++      +G+
8099  159 -----APQHRRgLLVSLNELMIV-----IGI 179

      ivPHiaRrLvGt.dhrwLLPaSAlLGAiLLllADllARtlfaPiElPvGi
      +  +i      h w      +++++ + +P G+
8099  180 LSAYISNYAFANvFHGW-----KYMFGGLVIPLGV 208

      vTAlIGaPyFl.....YLLrr<-*
      + A+  a+yFl+++++L+++
8099  209 LQAI--AMYFLppsprFLVMK      227

```

sugar_tr: domain 1 of 1, from 43 to 564: score 318.2, E = 9.6e-92

```

*->valvaalgGgflfGyDtgvigglalidflfrfgiltssgalaelvg
      +++++ G +l Gy +g+i+g+l +i+      l s+  ++
8099  43  SSVTAAVSG-LLVGYELGIISGALLQIK---TLLALSCHQE---- 80

```

FIGURE 3A


```

8099 213 ----AMYFLP-----PSPRFLVMKGQEGAASKVLGRLRALSDTTEELTV 252
      lpkasilkledakaersvdsLlsskSvgerdksqslsekqksqasgrpsss
8099 253 I----- 253
      atavqlvllrsrlekadlplkrvrvsrrrvlSkVSaeSgtdgersSgylN
      +s+l
8099 254 -----KSSL----- 257
      rkdvFYtGsisNvaeftkedpdkYrssslhgtrttvgnaesqstlrlddsr
8099 - ----- -
      esgdgdsssedlsektrgdggkessskeiretikllDfsvlknrtFl1
      k ++ +++++ ++ + + + r ++
8099 258 -----KDEYQYSFWDLFR---SKDNMRTRIMIG 282
      yaisnlfaslGffvPlvfLvsYaikslgldekeAsfLlsi.iGvsnivGR
      +++ +++ G l++ + ks g ++eA+ L s+++Gv+ +++
8099 283 LTLVFFVQITGQPNILFYASTVL-KSVGFQSNEAASLASTgVGVVKVIST 331
      pifGlvADkkgvrpTarhivyifnlsllalGlttlacPlatsfwgLvvyg
      + ++l+ D+ g ++ + s+++a+ l t+ + + ++ +c
8099 332 IPATLLVDHVGSKT-----FLCIGSSVMAASLVTMGIVNLNIHMNFTHIC 376
      ilFGfs.....
      +++++ ++ +++++ +++++ +++++ +++++ + +++
8099 377 RSHNSInqsldeviypgnlstnnntlrdhfkjsshrsslmplrndv 426
      .....
      ++++++ + + + +++++ + +++++ + + + +
8099 427 dkrgettsasllnaglshteyqivtdpgdvpafkwlslasllvyvaafs 476
      .....iGsygaLtfvvLvdLvG
      + ++ + ++ +++ +++ + + +i ++Ltf +++dL+g
8099 477 iglgpmpwlvlseifpggirgramaltssmnwgINLLISLTFITVTDLIG 526
      WlekfsnAfGl1lllfeGvavLvGPPiaGlLvDakttgdYtvaFyfsGill
      + + +++++ +a+L ++ + + t+ + ++ +
8099 527 -----LPWVCFIYTIMSLASLLFVVMF---IPE--TKGCSLEQIS---M 562
      llsgl<-*
      l+ +
8099 563 ELAKV 567

```

FIGURE 3C

USTAL W (1.74) multiple sequence alignment

```

h8099FL      MVPVENTEGPSLLNQKGTAVETEGSGSRHPPWARGCGMFTFLSSVTA AVSGLLVGYELG
37021|GALP    --TRANS--P--RTERMPDAKKQG-----RSNKAMTFFVCFLAALAGLLFGLDIG
               .  *:  *      .:: .  .:::*          *.      :*:  ..  *:*:***.*  :*:

h8099FL      ISGALLQIKTLLALSCHEQEMVVSSLVIGALLASLTGGVLIDRYGRRTAIILSSCLLGL
7021|GALP    IAGALPFI ADEFQITSHTQEWVVSSMMFGAAVGAVGSGWLSFKLGRKKSMLMIGAILFVA
*:***  *      :  ::.*  **  *****::**  ::::  .*  *      :  *:*:*****:  *:

h8099FL      SLVLILSLSYTVLIVGRIAIGVSI SLSSIATCVYIAEIA PQHRRGLLVSLNELMIVIGI
7021|GALP    SLFSAAAPNVEVLILSRVLLGLAVGVASYTAPLYLSEIAPEKIRGSMISMYQLMITIGI
**.*      :  .  ***:.*:  *:*****:  :  :  *:*****:  **  :*:  :***.***

h8099FL      SAYISNYAFANVFHGWKYMFGGLVIPLGVLQAIAMYFLPPSPRFLVMKGOEGAASKVLGR
7021|GALP    GAYLSDTAFS-YTGAWRWMLGVIIIPAILLLIGVFFLPDSPRWFAAKRRFVDAERVLLR
.***:.*:  *:      .*:*:*:*:  .:*  *.:***  ***:..  *      :  *.:**  *

h8099FL      RALSDTTE-ELTVIKSSLKDEYQYSFWDLFRSKDNMRTRIMIGLTLVFFVQITGQPNIL
7021|GALP    RDTSAEAKRELDEIRESLQVK-Q-SGWALFKENS NFRAVFLGVLLQVMQQTGMNVIM
*  *      :  :  **  *:***:  :  *  *  *  *:***:  :  :  :  *:  *  :  *:***  *:

h8099FL      YASTVLKSVGFSNEAASLASTG VGVVKVISTIPATLLVDHVGSKTFLCIGSSVMAASL
7021|GALP    YAPKIFELAGYTNTEQMWGTVIVGLTNVLATFIAIGLVDRWGRKPTLTGLFLVMAAG-
**.:***:  .*:  ..      .:.  ***:*.***:  *  ***:  *  *  *  :*  *****.

h8099FL      TMGIVNLNIHMNFTHICRSHNSINQSLDESVIYGPGNLSTNNNTLRDHFKGISHSRSSI
7021|GALP    -MGVLGTMMHI-----GI--HS--
*:*:..  *:      **  **

h8099FL      MPLRNDVDKRGGETTSASLLNAGLSHTEYQIVTDPGDVPAFLKWLSLASLLVYVAAFSIGI
7021|GALP    -P-----SA-----QYFAIAMLLMFIVGFAMS/
*          **      :  :  :  *  *:***:  :  :  :

h8099FL      GPMPWLVLSEIFPGGIRGRAMALTSSMNWGINLLISLTFLTVDLIGLPWVCFIYTIMSI
7021|GALP    GPLIWVLCSEIQPLKGRDFGITCSTATNWIANMIVGATFLTMLNTLGNANTFWVYAALN
**:*:  *:  ***  *      *.  .::  :::  **  *:***:  :  :  *  .  :*:  :.

h8099FL      ASLLFVVMFIPETKGCSELEQISMELAKVNYVKNNICFMSHHQEELVPKQPQKRKPQEQLI
7021|GALP    LFILLTLWLVPETKHSLEHIERNLMKGRKLR---EIGAHD-----
:***:  :*:***  ***:*.  *  *  .  :  :  :  :  :*.

h8099FL      ECNKL CGRGQSRQLSPET
7021|GALP    -----

```

FIGURE 4

USTAL W (1.74) multiple sequence alignment

```

18099FL      MVPVENTEGPSLLNQKGTAVETEGSGSRHPPWARGCG-MFTFLSSVTA AVSGLLVGYELG
9830|ARAE    --TRANS--P--RTERMVTINTESALT--PRSLRDTRRMNMFVS-VAAAVAGLLFGLDIG
      .  *:  *      .:: ::::***.: :  *  *  *  *: * :***:***.* ::*

18099FL      IISGALLQIKTLLALSCHQEMVVSLSLVIGALLASLTGGVLIDRYGRRTAIILSSCLLGL
9830|ARAE    VIAGALPFITDHFVLTSLRQEWVSSMMLGAAIGALFNGWLSFRLGRKYSLMAGAILFVL
      :*:***  *  . :*:..: ** *****:*** :*: *  *  *  *: :*: :*: *

18099FL      GSLVLILSLSYTVLIVGRIAIGVSISSLSIATCVYIAEIAPOHRRGLLVSLNELMIVIGI
830|ARAE     GSIGSAFATSVEMLIAARVVLGIAVGIASYTAPLYLSEMASENVRGKMISMYQLMVTLGI
      **:      : :  *  :*:..*.:*:~::~* : : :*:~*:~*:~* :*: :*: :*:~*:~*

8099FL      LSAYISNYAFANVFHG-WKYMFG-LVIPLGVLQAIAMYFLPPSPRFLVMKGQEGAASKVL
830|ARAE     VLAFLSDTAFS--YSGNWRAMLGVLALP-AVLLIILVVFLPNSPRWLAEKGRHIEAEVVL
      : :~*:~* :* : * : * : * : * : * : * : * : * : * : * : * : * : *

8099FL      GRLRALSDTT-EELTVIKSSLKDEYQYSFWDLFERSKDNMRTRIMIGLTLVFFVQITGQPN
830|ARAE     RMLRDTSEKAREELNEIRESLKLK-Q-GGWALFKINRNVRRVFLGMLLQAMQQFTGMNI
      **  *~*:~* :*:~*~* :  *  *  * : * : * : * : * : * : * : * : * : *

8099FL      ILFYASTVLKSVGFQSNEAASLASTGVGVVKVISTIPATLLVDHVGSKTFLCIGSSVMAA
830|ARAE     IMYYAPRIFKMAGFTTTEQQMIATLVVGLTFMFATFIAVFTVDKAGRKPALKIGFSVMAL
      *~*:~* . :* .** :.* :~* :~* :~* :~* :~* :~* :~* :~* :~* :~* :~*

8099FL      SLVTMGIVNLNIHMNFTHICRSHNSINQSLDESVIYGPGNLSTNNNTLRDHFKGISSHSR
830|ARAE     GTLVLG-----YC-----
      . :~*:~*      *

3099FL      SSLMPLRNDVDKRGETTSASLLNAGLSHTEYQIVTDPGDVPAFLKWLSLASLLVYVAAFS
330|ARAE     --LMQFDN-----G-TASS-----GLS-----WLSVGMTMMCIAGYA
      ** :  *      * *~*:~*      ***      ***~* . : :~*~*~*

3099FL      IGLGPMPWLVLSEIFPGGIRGRAMALTSSMNWGINLLISLTFLTVDLIGLPWVCFIYTI
330|ARAE     MSAAPVVWILCSEIQPLKCRDFGITCSTTNWVSNMIIGATFLTLLDSIGAAGTFWLYTA
      :. .*: :~*:~* *  *  . :~* :~* :~* :~* :~* :~* :~* :~* :~* :~*

3099FL      MSLASLLFVVMFIPETKGCSEIQISMELAKVNYVKNNICFMSHHQEELVPKQPQKRKPQE
330|ARAE     LNIAFVGITFWLIPETKNVTLEHIERKLMAGEKLRN-IGV-----
      :~*:~* : :~* :~*~*~* :~*~*~* :~* : :~*~*~*

099FL      QLLECNKLCGRGQSRQLSPET
30|ARAE     -----

```

FIGURE 5

09830 | ARAE
37021 | GALP
bh8099FL

-----TRANSPRTERMVTINTESALT--PRSLRDTRRMNMFVSVAAAVAGLLFGLDIGV
-----TRANSPRTERMPDAKKQG-----RSNKAMTFFVCFLAALAGLLFGLDIGV
MVPVENTEGPSLLNQKGTAVETEGSGSRHPPWARGCGMFTFLSSVTAAVSGLLVGYELGI
.. :.: :.: * :.: . **::***.* :.:

09830 | ARAE
37021 | GALP
bh8099FL

IAGALPFITDHFVLTSRLQEWVVSMMLGAAIGALFNGWLSFRLGRKYSLMAGAILFVLC
IAGALPFIADFEQITSHTEQEWVVSMMFGAAGVAVGSGWLSFKLGRKKSMLMIGAILFVAC
ISGALLQIKTLLALSCHQEEMVVSLLVIGALLASLTGGVLIDRYGRRTAILSSCLLGLC
*::** * : :.: ** *****:.* :.: . * * : ** : :.: . : * : *

9830 | ARAE
7021 | GALP
bh8099FL

SIGSAFATSVEMLIAARVVLGIAVGIASYTAPLYLSEMASENVRGKMISMYQLMVTGLGIV
SLFSAAAPNVEVLILSRVLLGLAVGVASYTAPLYLSEIAPEKIRGSMISMYQLMITIGII
SLVLILSLSYTVLIVGRIAIGVSISSSIATCVYIAEIAPOHRRGLLVSLNELMIVIGII
* : : . : ** . * : : : : : * : : : : * : : ** : : * : : * : : * : :

9830 | ARAE
7021 | GALP
bh8099FL

LAFSLDTAFS--YSGNWRAMLGVLP--AVLLIILVVFLPNSPRWLAEKGRHIEAEEVLR
GAYLSDTAFS--YTGAWRWMLGVIIIP--AILLLIGVFFLPDSPRWFAAKRRFVDAERVLL
SAYISNYAFANVFHG--WKYMFG--LVIPLGVLQAIAMYFLPPSPRFLVMKGQEGAASKVLG
::: ** : * * : * : * : : * . : * : * * * : . * : * . . . *

9830 | ARAE
7021 | GALP
bh8099FL

MLRDTSEKAREELNEIRESLKLK-Q-GGWALFKINRNVRAVFLGMLLQAMQOFTGMNII
RLRDTSAEAKRELDEIRESLQVK-Q-SGWALFKENSNFRAVFLGVLLQVMQOFTGMNVI
RLRALSDDT--EELTVIKSSLKDEYQYSFWDLFKSKDNMRTRIMIGLTLVFFVQITGQPNII
** * : . ** * : . : * : * * : : * . * : : : : * : * : * *

9830 | ARAE
021 | GALP
bh8099FL

MYAPRIFKMGFTTTEQOMIATLVVGLTFMFATFIAVFTVDKAGRKPALKIGFSVMALG
MYAPKIFELAGYTNTTEQMWGTIVVGLTNVLFATFIAIGLVDRWGRKPTTLTGLFLVMAAG
LFYASTVLKSVGFQSNAAASLASTGVGVVKVISTIPATLLVDHVGSKTFLCIGSSVMAAS
: : * . : : . * : . . : : * : . : : : * * : * * . * : * * * .

830 | ARAE
021 | GALP
bh8099FL

TLVLGYCLMQFDN-----GTASSG--
MGVLG-TMMHI-----GIHSPS--
LVTMGIVNLNIHMNFTHICRSHNSINQSLDESVIYGPGLNSTNNNTLRDHFKGISHSRS
..* :.: * *

830 | ARAE
021 | GALP
bh8099FL

-----LSWLSVGMTTMMCIAGYAM
-----AQYFAIAMLLMFIVGFAM
SLMPLRNDVDKRGGETTSASLLNAGLSHTEYQIVTDPGDVPAFLKWLSLASLLVYVAAFSI
.: : . : : : . : : : .

830 | ARAE
021 | GALP
bh8099FL

SAAPVVWILCSEIQP--LKCRDFGITCSTTTNWVSNMIIIGATFLTLTLLDSIGAAGTFWLYT
SAGPLIWVLCSEIQP--LKGRDFGITCSTATNWIANMIVGATFLTMLNTLGNANTFWVYA
GLGPMPLVLSEIFPGGIRGRAMALTSS--MNWGINLLISLTFLTVTDLIGLPWVCFIYT
. . * : * : * * * : : * : : * . * * : : . * : . : : *

30 | ARAE
21 | GALP
bh8099FL

ALNIAFVGITFWLIPETKNVTLEHIERKLMAGEKLRN----IGV-----
ALNVLFILLTLWLVPETKHSLEHIERNLMKGRKLRE----IGAHD-----
IMSLASLLFVVMFIPETKGCSEIQISMELAKVNVKNNICFMSHHQEELVPKQPQKRKPQ
.: : : . : : * * * : * : * . : * : : .

30 | ARAE
21 | GALP
bh8099FL

EQLLECNKLCGRGQSRQLSPET

FIGURE 6

CLUSTAL W (1.74) multiple sequence alignment

```

bh8099FL      MVPVENTEGPSLLNQKGTAVETEGSG---SRHPPWARG-CGMFTFLSSVTA AVSGLLVGY
02168Patent    ---MTSDHEHMTAVCASHVQTHGSQ LQIQKLSPCFRPPTPAFRIS SSIILLGAG-LAGP
               .::      :      .:  *:*.*      .:  .*  *      *  :  **:      :*  *.*

bh8099FL      ELGII SGALLQIKTLLALSCH EQEMVSSLVIGALLASLTGGVLIDRYGRRTAIILSSCL
02168Patent    STGDRWFGVSVVGTGLFLPPLQLLLPRL LFTHAILERLHLWLALPPVLVLGHALLH-CK
               .  *      .:  :  *  *  *  .  :  :      *  .  *:  *  *      :  :      :*  *

bh8099FL      LGLGSLVLILSLSYTVLIVGRIAIGVSI SLSSIATCVYIAEIAPOH--RRGLLVSLNELM
02168Patent    VGGSTARAGDQLVQRVLL-IVFLHRWVQVWPZGTEVDILGMGSRTGGRRGP ELRP---G
               :*  .:      .*      **:  :  :  :  :  .  .  *  *  *      :  .:      ***  :

bh8099FL      IVIGILSAYISNYAFANVFHGWKYMFG LVIPLGVLQAIAMYFLPPSPRFLVMKGQEGAAS
02168Patent    FRISILSAYISNYAFANVFHGWKYMFG LVIPLGVLQAIAMYFLPPSPRFLVMKGQEGAAS
               :  *  .*****

bh8099FL      KVLGRLRALSDTTEELTVIKSSLKDEYQYSFWD LFRSKDNMRTRIMIGLTLVFFVQITGQ
02168Patent    KVLGRLRALSDTTEELTVIKSSLKDEYQYSFWD LFRSKDNMRTRIMIGLTLVFFVQITGQ
               *****

bh8099FL      PNILFYASTVLKSVGFQ SNEAASLASTG VGVVKVISTIPATLLVDHVGSKTFLCIGSSVM
02168Patent    PNILFYASTVLKSVGFQ SNEAASLASTG VGVVKVISTIPATLLVDHVGSKTFLCIG---
               *****

bh8099FL      AASLVTMGIVNLNIHMNFTHICRSHNSINQSLDES VIYGPGNLSTNNNTLRDHFKGIS SH
02168Patent    -----

bh8099FL      SRSSLMPLRNDVDKRGETTSASLLNAGLSHTEYQ IVTDPGDVPAFLKWLSLASLLVYVAA
02168Patent    -----LLNAGLSHTEYQ IVTDPGDVPAFLKWLSLASLLVYVAA
               *****

bh8099FL      FSI GLGMPWLVLSEIFPGGIRGRAMALTSSMNWGINLLISL TFLTVDLIGLPWVCFIY
02168Patent    FSI GLGMPWLVLSEIFPGGIRGRAMALTSSMNWGINLLISL TFLTVDLIGLPWVCFIY
               *****

bh8099FL      TIMSLASLLFVVMFI PETKGCSLEQISMELAKVNYVKNNI CFMSHHQEELVPKQPQKRKP
02168Patent    TIMSLASLLFVVMFI PETKGCSLEQISMELAKVNYVKNNI CFMSHHQEELVPKQPQKRKP
               *****

bh8099FL      QEQLLECNKLCGRGQSRQLSPET
02168Patent    QEQLLECNKLCGRGQSRQLSPET
               *****

```

FIGURE 7

Input file Fbh46455FL.seq; Output File Fbh46455FL.tra
Sequence length 2230

```

GTCGACCCACGCGTCCGGCAACATGGCGGCTGCCGTGGTGCAGCGCCCGGGCTGAGCGACAGCAAGTGCAGCGGGCTCC
TACCCCGGGTGAGGGGTGGCCTCCGCGTGGGATCGTGCCCTCTTCAGCCCGCTCCTGTCCCCGACATCACGTGTATTCC
GCACGTCCCCTCCGCGCTGTGTGTCTACTGAGACGGGGAGGCGTGACAGGGCCCGGGTCCCTTCTCAGTGGTGTCTGT
GCTTCAGGGCAAGCTCCCCGTCTCCGGGCGCACTTCCCTCGCCTGTGTTCCGGTCCATCCTCCTTTCTCCAGCCTCCTCC

                                M  A  G  S  D
CCTCGCAGGTGGGATCGTCGGTGGGACCGGAGCGCGGGCGGGCGCGGCCCGGGACC ATG GCC GGG TCC GAC      5
                                15

T  A  P  F  L  S  Q  A  D  D  P  D  D  G  P  V  P  G  T  P      25
ACC GCG CCC TTC CTC AGC CAG GCG GAT GAC CCG GAC GAC GGG CCA GTG CCT GGC ACC CCG      75

G  L  P  G  S  T  G  N  P  K  S  E  E  P  E  V  P  D  Q  E      45
GGG TTG CCA GGG TCC ACG GGG AAC CCG AAG TCC GAG GAG CCC GAG GTC CCG GAC CAG GAG      135

G  L  Q  R  I  T  G  L  S  P  G  R  S  A  L  I  V  A  V  L      65
GGG CTG CAG CGC ATC ACC GGC CTG TCT CCC GGC CGT TCG GCT CTC ATA GTG GCG GTG CTG      195

C  Y  I  N  L  L  N  Y  M  D  R  F  T  V  A  G  V  L  P  D      85
TGC TAC ATC AAT CTC CTG AAC TAC ATG GAC CGC TTC ACC GTG GCT GGC GTC CTT CCC GAC      255

L  E  Q  F  F  N  I  G  D  S  S  S  G  L  I  Q  T  V  F  I      105
ATC GAG CAG TTC TTC AAC ATC GGG GAC AGT AGC TCT GGG CTC ATC CAG ACC GTG TTC ATC      315

S  S  Y  M  V  L  A  P  V  F  G  Y  L  G  D  R  Y  N  R  K      125
TCC AGT TAC ATG GTG TTG GCA CCT GTG TTT GGC TAC CTG GGT GAC AGG TAC AAT CGG AAG      375

Y  L  M  C  G  G  I  A  F  W  S  L  V  T  L  G  S  S  F  I      145
TAT CTC ATG TGC GGG GGC ATT GCC TTC TGG TCC CTG GTG ACA CTG GGG TCA TCC TTC ATC      435

P  G  E  H  F  W  L  L  L  L  T  R  G  L  V  G  V  G  E  A      165
CCC GGA GAG CAT TTC TGG CTG CTC CTC CTG ACC CGG GGC CTG GTG GGG GTC GGG GAG GCC      495

S  Y  S  T  I  A  P  T  L  I  A  D  L  F  V  A  D  Q  R  S      185
AGT TAT TCC ACC ATC GCG CCC ACT CTC ATT GCC GAC CTC TTT GTG GCC GAC CAG CGG AGC      555

R  M  L  S  I  F  Y  F  A  I  P  V  G  S  G  L  G  Y  I  A      205
CGG ATG CTC AGC ATC TTC TAC TTT GCC ATT CCG GTG GGC AGT GGT CTG GGC TAC ATT GCA      615

G  S  K  V  K  D  M  A  G  D  W  H  W  A  L  R  V  T  P  G      225
GGC TCC AAA GTG AAG GAT ATG GCT GGA GAC TGG CAC TGG GCT CTG AGG GTG ACA CCG GGT      675

L  G  V  V  A  V  L  L  L  F  L  V  V  R  E  P  P  R  G  A      245
CTA GGA GTG GTG GCC GTT CTG CTG CTG TTC CTG GTA GTG CGG GAG CCG CCA AGG GGA GCC      735

V  E  R  H  S  D  L  P  P  L  N  P  T  S  W  W  A  D  L  R      265
GTG GAG CGC CAC TCA GAT TTG CCA CCC CTG AAC CCC ACC TCG TGG TGG GCA GAT CTG AGG      795

A  L  A  R  N  P  S  F  V  L  S  S  L  G  F  T  A  V  A  F      285
GCT CTG GCA AGA AAT CCT AGT TTC GTC CTG TCT TCC CTG GGC TTC ACT GCT GTG GCC TTT      855

V  T  G  S  L  A  L  W  A  P  A  F  L  L  R  S  R  V  V  L      305
GTC ACG GGC TCC CTG GCT CTG TGG GCT CCG GCA TTC CTG CTG CGT TCC CGC GTG GTC CTT      915

```

FIGURE 8A

G	E	T	P	P	C	L	P	G	D	S	C	S	S	S	D	S	L	I	F	325
GGG	GAG	ACC	CCA	CCC	TGC	CTT	CCC	GGA	GAC	TCC	TGC	TCT	TCC	TCT	GAC	AGT	CTC	ATC	TTT	975
G	L	I	T	C	L	T	G	V	L	G	V	G	L	G	V	E	I	S	R	345
GGA	CTC	ATC	ACC	TGC	CTG	ACC	GGA	GTC	CTG	GGT	GTG	GGC	CTG	GGT	GTG	GAG	ATC	AGC	CGC	1035
R	L	R	H	S	N	P	R	A	D	P	L	V	C	A	T	G	L	L	G	365
CGG	CTC	CGC	CAC	TCC	AAC	CCC	CGG	GCT	GAT	CCC	CTG	GTC	TGT	GCC	ACT	GGC	CTC	CTG	GGC	1095
S	A	P	F	L	F	L	S	L	A	C	A	R	G	S	I	V	A	T	Y	385
TCT	GCA	CCC	TTC	CTC	TTC	CTG	TCC	CTT	GCC	TGC	GCC	CGT	GGT	AGC	ATC	GTG	GCC	ACT	TAT	1155
I	F	I	F	I	G	E	T	L	L	S	M	N	W	A	I	V	A	D	I	405
ATT	TTC	ATC	TTC	ATT	GGA	GAG	ACC	CTC	CTG	TCC	ATG	AAC	TGG	GCC	ATC	GTG	GCC	GAC	ATT	1215
L	L	Y	V	V	I	P	T	R	R	S	T	A	E	A	F	Q	I	V	L	425
CTG	CTG	TAC	GTG	GTG	ATC	CCT	ACC	CGA	CGC	TCC	ACC	GCC	GAG	GCC	TTC	CAG	ATC	GTG	CTG	1275
S	H	L	L	G	D	A	G	S	P	Y	L	I	G	L	I	S	D	R	L	445
TCC	CAC	CTG	CTG	GGT	GAT	GCT	GGG	AGC	CCC	TAC	CTC	ATT	GGC	CTG	ATC	TCT	GAC	CGC	CTG	1335
R	R	N	W	P	P	S	F	L	S	E	F	R	A	L	Q	F	S	L	M	465
CGC	CGG	AAC	TGG	CCC	CCC	TCC	TTC	TTG	TCC	GAG	TTC	CGG	GCT	CTG	CAG	TTC	TCG	CTC	ATG	1395
F	C	A	F	V	G	A	L	G	G	A	A	F	L	G	T	A	I	F	I	485
CTC	TGC	GCG	TTT	GTT	GGG	GCA	CTG	GGC	GGC	GCA	GCC	TTC	CTG	GGC	ACC	GCC	ATC	TTC	ATT	1455
E	A	D	R	R	R	A	Q	L	H	V	Q	G	L	L	H	E	A	G	S	505
GAG	GCC	GAC	CGC	CGG	CGG	GCA	CAG	CTG	CAC	GTG	CAG	GGC	CTG	CTG	CAC	GAA	GCA	GGG	TCC	1515
T	D	D	R	I	V	V	P	Q	R	G	R	S	T	R	V	P	V	A	S	525
ACA	GAC	GAC	CGG	ATT	GTG	GTG	CCC	CAG	CGG	GGC	CGC	TCC	ACC	CGC	GTG	CCC	GTG	GCC	AGT	1575
V	L	I	*																	529
GTG	CTC	ATC	TGA																	1587

GAGGC'TGCCGCTCACCTACCTGCACATCTGCCACAGCTGGCCCTGGGCCCCACCCACGAAGGGCCTGGGCCTAACCCCT
 TGGCCTGGCCCAGCTTCCAGAGGGACCCTGGGCCGTGTGCCAGCTCCCAGACACTACATGGGTAGCTCAGGGGAGGAGG
 TGGGGGTCCAGGAGGGGGATCCCTCTCCACAGGGGCAGCCCCAAGGGCTCGGTGCTATTTGTAAACGGAATAAAATTTGT
 AGCCAGAAAAAAAAAAAAAAAAAGGGCGGCCGC

FIGURE 8B

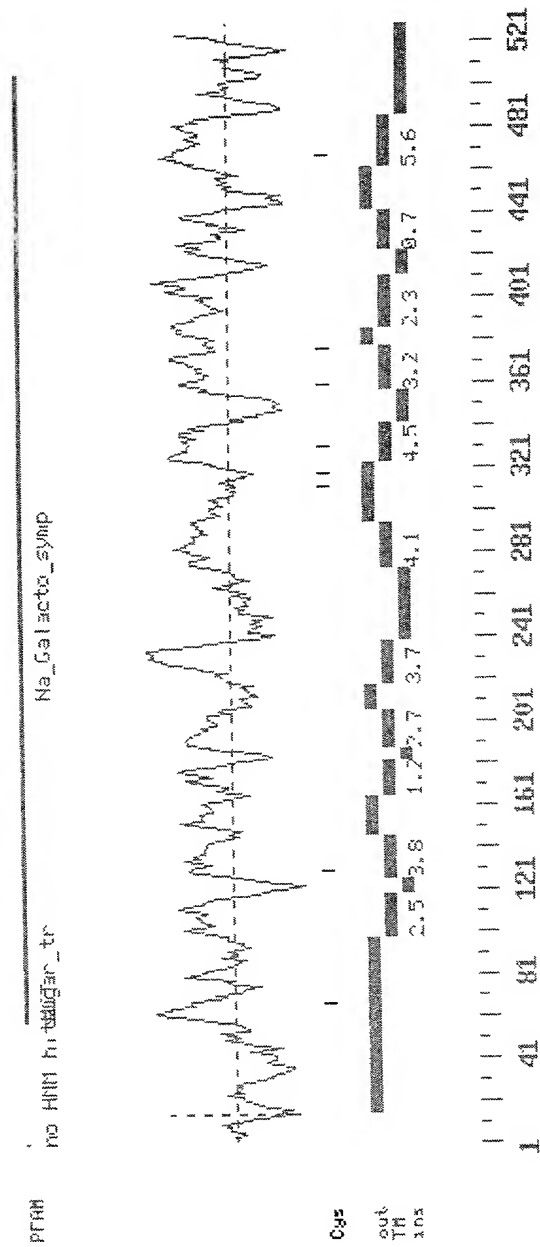


FIGURE 9

Protein Family / Domain Matches, HMMer version 2

Searching for complete domains in PFAM

hmmpfam - search a single seq against HMM database

HMMER 2.1.1 (Dec 1998)

Copyright (C) 1992-1998 Washington University School of Medicine

HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /prod/ddm/seqanal/PFAM/pfam6.4/Pfam
Sequence file: /prod/ddm/wspace/orfanal/oa-script.9015.seq

Query: 46455

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
sugar_tr 1	Sugar (and other) transporter	-63.4	0.00016	
Na_Galacto_symp 1	Sodium:galactoside symporter family	-121.2	0.17	
MCT 1	Monocarboxylate transporter	-208.2	0.32	

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
MCT	1/1	60	473	..	1 611 []	-208.2	0.32
sugar_tr	1/1	58	487	..	1 488 []	-63.4	0.00016
Na_Galacto_symp	1/1	212	505	..	1 285 []	-121.2	0.17

Alignments of top-scoring domains:

MCT: domain 1 of 1, from 60 to 473: score -208.2, E = 0.32

```

*->kpPDGGwGWvVVFasFlingfvdGfiksfgvffsellqeetlfnesk
      ++V      in  +  +++ ++ ++  e++fn+++
46455  60  -----LIVAVLCYINLLNYMDRFTVAGVLPDI---EQFFNIGD 94

      sdvdtAwIgsImlavllfsGPlsSilvnrfGcRivmiaGgllagaGllla
      s      ++I  ++++ +  ++P+ + l +r+  ++ m+ G ++ ++  l +
46455  95  SS--SGLIQTTFISSYMLAPVFGYLGDRYNRKYLMCGGIAFWSLVTLGS 142

      sFst..niwelyltfGvitGlgfifqPai.vilgqYFe.....KrRsl
      sF ++++ w+l lt G ++G+G + ++  + ++ ++  F  +++++ +s+
46455  143 SFIPgeHFWLLLLTRG-LVGVGGEA-SYSTIAPtLIADLFVadqrsRMLSI 190

      AtGiAvaGsGvGtvvppllqflidnyGsDWrgallilggillncvicGa
      +  GsG+G      +++ ++ d  G DW +al+++ g+  + v++ +
46455  191 FYFAIPVGSGLGY----IAGSKVKDMAG-DWHWALRVTPGLGVVAVLLLF 235

      lllRplepsvpqdekdketlkeakkkkendtettkeetepklslpkas
      l +R+      +p++ ++
46455  236 LVVRE----PPRGAVER----- 248

      ilkledakaersvdsLlsskSvgerdksqkseqksgasgrpsssatavq

46455  - -----

      lvllrsrlekadlplkrvrsvrrrvlSkVSaeSgtdgersSgylNrkdvF
      +s+l
46455  249 ----HSDL----- 252

      YtGsisNvaeekedpdkYrssslhgttrttvgnaesqstlrlldsresgdg

46455  - -----

```

FIGURE 10A

```

dsssedlsektrgdggkessskeiretikllDfsvlk.nrtFl1....
          ++++ ++ D++ l +n++F+l++ +
46455 253 -----PPLNPTSWWADLRALARNPSPFVLsslg 279

.yaisnlfaslGffvPlvfLvsYaikslg.....ldekeAsflls
a++++ sl++ +P++ L s++ lg++++ ++++ ++++ s++
46455 280 fTAVAFVTGSLALWAPAFLLRSRV--VLGetppclpgdsCSSSD-SLIFG 326

iiGvsnivGRpifGlvADkkgvrpTarhivyifnlsllal..GlttlacP
i + v +++G+ + +++ r + ++ ll ++l++ ++
46455 327 LITCLTGVLGVLGVEISRRLRHSNPRADPLVCATGLLGSapFLFLSLAC 376

latsfwgLvvy CilFGfs.iGsygaLtfvvLvdlvgWlek.....fsnA
+s++ +++ i+ G + +++a+ + +L v + ++++ + f+
46455 377 ARGSIYATYIF-IFIGETLLSMNWAIVADILLYVVI-PTRrstaefQIV 424

fGllllfeGvavLvGPPi...aGlLvDakttdgYtvaFyfsGillllsg
++ ll G + L+G ++++ + ++++ + ++ ++ ++++
46455 425 LSHLLGDAGSPYLI GLISdrLrRNWPPSF--LSEFRALQFSLMLCAFGA 472

l<-*
l
46455 473 L 473

sugar_tr: domain 1 of 1, from 58 to 487: score -63.4, E = 0.00016
*->valvaalgGgflfGyDtgvggflalidflfrfglltssgalaelvg
al++a+ + + + ++++ ++ + f+ + +s+
46455 58 SALIVAVLC-YINLLNYMDRFTVAGVLPDIEQFFNIGDSS----- 96

ystvltglvvsiffllGrliGslfaGklgdrfGRkksllialvlvfiGall
+gl+ ++f+ ++ + ++G+lgdr+ Rk+ + ++++ + +l+
46455 97 -----SGLIQTVFISSYMLAPVFGYLGDRYNRKYLMCGGIAFWSLVTLG 141

sgaapgytTiGlwafyllivGRvlvGlgvGgasvlvPmYisEiAPkalRG
s ++pg +f+ll++ R lvG g s ++P++i+ + R
46455 142 SSFIPGE-----HFWLLLLTRGLVGVGEASYSTIAPTILIADLFVADQRS 185

algslyqlaitiGilvAaiiglglnktnndsalsnswgWRiplglqlvpal
++ s++ +ai +G +++i g ++ +++d +w R+ gl+ v l
46455 186 RMLSIFYFAIPVGSGLGYIAGSKVKDMAGDW---HWALRVTPGLGVVAVL 232

llligllflPESPRwLvekgleeArevLaklrgvedvdqeieikaee
ll++++ P rg + + + + + +++
46455 233 LLFLVVREPP-----RGAVERHSDLPLPLNPTSW 260

atvseekagkaswgelfrgrtrpkvrqrllmgvmlqafqQltGiNaifYY
+ + + l r+++ +l + + +a+ +tG ++ +
46455 261 WA-----DLRALARNPS-----FVLSSLGFTAVAFVTG--SLALW 293

sptifks.....vGvsdsvasllvtiivgvvNfvfTfvaLiflvD
+p ++ ++ +++++ +ds +s ++i+g+++ ++ + + + l
46455 294 APAFLLRsrvvlggetppCLPGDS-CSSSDSLIFGLITCLTGVLG-VGLGV 341

rfGRR.....pllllGaagmaicflilgasigvalllllnkpkdpss
+ RR +++++ +pl++ ++ ++ fl+l+ l++ ++
46455 342 EISRRLrhsnpradPLVCATGLLGSAPFLFLS-----LACARGs----- 380

kaagivaivfillfiafFalgwGpipwilsElFPtkv...Rskalala
iv++++fi+ + + + w+i++++ v +++Rs+a a+
46455 381 -----IVATYIFIFIGE-TLLSMNWAIVADILLYVViPtrRSTAEAFQ 422

taanwlanfiigflfpyitgaigl....algyvflvfagl...lvlfil
++ l + + + py+ g i+ + +++++++ f +l+ +l l+ +
46455 423 IVLSHLLGDAGS---PYLIGLISdrLrrNWPPSFLEFRALqfsLMLCAF 469

```

FIGURE 10B

```

fvfffvPETkGrLLEieelf<-*
++ ++ ++ G ++i
46455 470 VGALGGAFLG---TAIFIEA 487

Na_Galacto_symp: domain 1 of 1, from 212 to 505: score -121.2, E = 0.17
*->qLG.yfffalV...LslagvllwiCf....fgtkEvySssdtreng
+ G+++++ V+++L++++vll+++ ++++g E+ sd ++ +
46455 212 MAGdWHWALRVtpgLGVVAVLLLFVreppRGAVRH--SDLPPLN 256

qkttslqslklakNdQ..LliLclaal fyllainilgg.aqlYYvtYv
++ ++l++la+N++ L L++ a + ++ +l ++a l + v
46455 257 PTSW--WADLRALARNPSfvLSSLGFTAVAFVTGSLALWApFLLRSRV 304

LG.dpelFs.....ylllynilvgligslLfPrLvkrf..gkktv
LG++p +++++ +++++ +++++ +l g g+ L ++r++++ ++
46455 305 LGeTPPCLPgdsccssdsIFGLITCLTGVLGVGLGVEISRRLrhSNPRA 354

FagcivlmvlgslLiFfvagsslal.ilvliflagilqqlvtllvWvlQV
+ ++lg ++ F++ +++ a + +v++ ++ + + +W++
46455 355 DPLVCATGLLG-SAPFLFLSLACARgSIVATYIFIFIGETLLSMNWAI-- 401

IMvsDtVDYGEwktGvRlEGlvvSvflfvKlGlAlsGalvGwiL..gyi
v+D+ Y t +R+++ ++ l l lG A s l+G+i ++
46455 402 --VADILLYVVIPT-RRSTAEAFQIVLSHL-LGDAGSPYLIGLISdrLRR 447

GYvanasqststalqQlvfilalFalPpallllaaafimlrfYkLtekkla
+ ++ s al+ f l l a++ al +a + ++ f+ + + +
46455 448 NWPPSFL-SEFRALQ---FSLMLCAFVGALGGAFLGTAIFIEADRRRAQ 493

eIveeLekWrtrkrk<-*
v L+ + +
46455 494 LHVQGLL---HEAGS 505

```

FIGURE 10C

CLUSTAL W (1.74) multiple sequence alignment

```

Fbh46455FL  MAGSDTAPFLSQADDPDDGPVPGTPGLPGSTGNPKSEEPEVPDQEGLQRITGLSP---G-
Z92825      MVRNKVAPVEDGANIQRFEP--P--PYTT--P-TDSPEDKIRSNSTATTASQPEFQGC
          * . . . * . . * :   :   * * * * : * * : . . * . . * *

Fbh46455FL  RSLIVAVLCYINLLNYMDRFTVAGVLPDIEQFFNIGDSSSLIQTTFISSYMLAPVFG
Z92825      WTIVVVAILFIINLLNYMDRYTIAGVLNDVQTYYNISDAWAGLIQTTFMVFFIIFSPICG
          : : * * : * * * * * * * * : : : * * : * * * * : : : : *

Fbh46455FL  YLGDRYNRKILMCGGIAFWSLVTLGSSFIIPGEHFWLLLLTRGLVGVGEASYSTIAPTIA
Z92825      FLGDRYNRKWIFVVGIAIWVSAVFASTFIPSNQFWLFLFRGIVGIGEASYAIISPTVIA
          : * * * * * : :   * * : * . . . : * * : : * * : * * : * * : *

Fbh46455FL  DLFVADQRSRMLSIIFYFAIPVGSGLGYIAGSKVKDMAGDWHWALRVTPGLGVVAVLLLFL
Z92825      DMFTGVLRSRMLMVFYFAIPFGCGLGFVVGSAVASWTGHWQWGVVRTGVLGIVCLLLIIV
          * : * . . * * * : * * * * . * * : . . * * * . : * * : * * : * : :

Fbh46455FL  VVREPPRGAVERHS-DLPPLNP-TSWWADLRALARNPFLVSSLGFTAVAFVTGSLALWA
Z92825      FVREPERGKAEREKGEIAASTEATSYLDDMKDLLSNATYVTSSLGYTATVFMVGTLAWWA
          . * * * * * * * . * . . . * :   * : : * * : : * * * * : * : * *

Fbh46455FL  PAFLLSRVVLGETPPCLPGDSCSSSDSLIFGLITCLTGVLGVGLGVEIS---RR---L
Z92825      PITIQYADSAR-RNGTITE-DQKANIN-LVFGALTCVGGVLGVAIGTLVSNMWSRGVGPFF
          * :   :   . . .   * . : . : * * * : * : * * * : * . : *   :

Fbh46455FL  RHSNP-RADELVCATGLLGSAPFLFLSLACARGSIVATYIFIFIGETLLSMNWAIVADIL
Z92825      KHIQTVRADALVCAIGAACIPTLILAIQNIESNMNFAWGMLFICIVASSFNWATNVDLL
          : * : . * * . * * * *   . * * : : :   . . . : : : * *   * : * *

Fbh46455FL  LYVVIPTRRSTAEAFQIVLSHLLGDAGSPYILIGLISDRLRRNWPPSFLSEFRALQFSLML
Z92825      LSVVVPQRRSSASSWQILISHMFGDASGPYILGLISDAIRGNED-TAQAHYKSLVTSFWL
          * * * : * * * : * . : : * * : * * : * * . * * :   : : : * * : *

Fbh46455FL  CAFVGALGGAFLGTAIFIEADRRRAQLHVQGLLHEAGSTDDRIVVPQGRGRSTRVPVASV
Z92825      CVGTLVLSVILFGISAITVVKDKARFNEIMLAQANKDNTSSG--TLPIEDRNTEDETGSE
          * . . * .   * : * * :   * : * :   : . : : . . . . : * . * . * . . *

Fbh46455FL  LI--
Z92825      VQHM
          :

```

FIGURE 11

CACGCGTCCGCCCACGCGTCCGCCACGCGTCCGAGCCCCCTTTCAAGCCTTAGCTTCCGGCTCCAAGCCGACCCCCTC
CCCCTCCCTGTCCCCTTCCCCTTCTCCCATCCCCTCTCTCGGCCACAGCGTCTTGTTAGTCTCTCCCTCTACTCCGCAA

FIGURE 12A

H	P	R	L	Q	D	Y	Y	V	V	I	L	C	P	T	E	M	D	V	Q	343
CAT	CCT	AGG	CTC	CAG	GAT	TAT	TAT	GTG	GTG	ATT	TTG	TGT	CCT	ACT	GAA	ATG	GAT	GTA	CAG	1029
V	R	R	V	L	Q	I	P	M	W	S	Q	R	V	I	Y	L	Q	G	S	363
GTT	CGA	AGG	GTA	CTG	CAG	ATT	CCA	ATG	TGG	TCC	CAA	CGA	GTT	ATC	TAC	CTT	CAA	GGT	TCA	1089
A	L	K	D	Q	D	L	L	R	A	K	M	D	D	A	E	A	C	F	I	383
GCC	CTT	AAA	GAT	CAA	GAC	CTA	TTG	AGA	GCA	AAG	ATG	GAT	GAC	GCT	GAG	GCC	TGT	TTT	ATT	1149
L	S	S	R	C	E	V	D	R	T	S	S	D	H	Q	T	I	L	R	A	403
CTC	AGT	AGC	CGT	TGT	GAA	GTG	GAT	AGG	ACA	TCA	TCT	GAT	CAC	CAA	ACA	ATT	TTG	AGA	GCA	1209
W	A	V	K	D	F	A	P	N	C	P	L	Y	V	Q	I	L	K	P	E	423
TGG	GCT	GTG	AAA	GAT	TTT	GCT	CCA	AAT	TGT	CCT	TTG	TAT	GTC	CAG	ATA	TTA	AAG	CCT	GAA	1269
N	K	F	H	I	K	F	A	D	H	V	V	C	E	E	E	F	K	Y	A	443
AAT	AAA	TTT	CAC	ATC	AAA	TTT	GCT	GAT	CAT	GTT	GTT	TGT	GAA	GAA	GAG	TTT	AAA	TAC	GCC	1329
M	L	A	L	N	C	I	C	P	A	T	S	T	L	I	T	L	L	V	H	463
ATG	TTA	GCT	TTA	AAC	TGT	ATA	TGC	CCA	GCA	ACA	TCT	ACA	CTT	ATT	ACA	CTA	CTG	GTT	CAT	1389
T	S	R	G	Q	E	G	Q	Q	S	P	E	Q	W	Q	K	M	Y	G	R	483
ACC	TCT	AGA	GGG	CAA	GAA	GGC	CAG	CAA	TCG	CCA	GAA	CAA	TGG	CAG	AAG	ATG	TAC	GGT	AGA	1449
C	S	G	N	E	V	Y	H	I	V	L	E	E	S	T	F	F	A	E	Y	503
TGC	TCC	GGG	AAT	GAA	GTC	TAC	CAC	ATT	GTT	TTG	GAA	GAA	AGT	ACA	TTT	TTT	GCT	GAA	TAT	1509
E	G	K	S	F	T	Y	A	S	F	H	A	H	K	K	F	G	V	C	L	523
GAA	GGA	AAG	AGT	TTT	ACA	TAT	GCC	TCT	TTC	CAT	GCA	CAC	AAA	AAG	TTT	GGC	GTC	TGC	TTG	1569
I	G	V	R	R	E	D	N	K	N	I	L	L	N	P	G	P	R	Y	I	543
ATT	GGT	GTT	AGG	AGG	GAG	GAT	AAT	AAA	AAC	ATT	TTG	CTG	AAT	CCA	GGT	CCT	CGA	TAC	ATT	1629
M	N	S	T	D	I	C	F	Y	I	N	I	T	K	E	E	N	S	A	F	563
ATG	AAT	TCT	ACG	GAC	ATA	TGC	TTT	TAT	ATT	AAT	ATT	ACC	AAA	GAA	GAG	AAT	TCA	GCA	TTT	1689
K	N	Q	D	Q	Q	R	K	S	N	V	S	R	S	F	Y	H	G	P	S	583
AAA	AAC	CAA	GAC	CAG	CAG	AGA	AAA	AGC	AAT	GTG	TCC	AGG	TCG	TTT	TAT	CAT	GGA	CCT	TCC	1749
R	L	P	V	H	S	I	I	A	S	M	G	T	V	A	I	D	L	Q	D	603
AGA	TTA	CCT	GTA	CAT	AGC	ATA	ATT	GCC	AGC	ATG	GGT	ACT	GTG	GCT	ATA	GAC	CTG	CAA	GAT	1809
T	S	C	R	S	A	S	G	P	T	L	S	L	P	T	E	G	S	K	E	623
ACA	AGC	TGT	AGA	TCA	GCA	AGT	GGC	CCT	ACC	CTG	TCT	CTT	CCT	ACA	GAG	GGA	AGC	AAA	GAA	1869
I	R	R	P	S	I	A	P	V	L	E	V	A	D	T	S	S	I	Q	T	643
ATA	AGA	AGA	CCT	AGC	ATT	GCT	CCT	GTT	TTA	GAG	GTT	GCA	GAT	ACA	TCA	TCG	ATT	CAA	ACA	1929
C	D	L	L	S	D	Q	S	E	D	E	T	T	P	D	E	E	M	S	S	663
TGT	GAT	CTT	CTA	AGT	GAC	CAA	TCA	GAA	GAT	GAA	ACT	ACA	CCA	GAT	GAA	GAA	ATG	TCT	TCA	1989
N	L	E	Y	A	K	G	Y	P	P	Y	S	P	Y	I	G	S	S	P	T	683
AAC	TTA	GAG	TAT	GCT	AAA	GGT	TAC	CCA	CCT	TAT	TCT	CCA	TAT	ATA	GGA	AGT	TCA	CCC	ACT	2049
F	C	H	L	L	H	E	K	V	P	F	C	C	L	R	L	D	K	S	C	703
TTT	TGT	CAT	CTC	CTT	CAT	GAA	AAA	GTA	CCA	TTT	TGC	TGC	TTA	AGA	TTA	GAC	AAG	AGT	TGC	2109

FIGURE 12B

Q	H	N	Y	Y	E	D	A	K	A	Y	G	F	K	N	K	L	I	I	V	723
CAA	CAT	AAC	TAC	TAT	GAG	GAT	GCA	AAA	GCC	TAT	GGA	TTC	AAA	AAT	AAA	CTA	ATT	ATA	GTT	2169
A	A	E	T	A	G	N	G	L	Y	N	F	I	V	P	L	R	A	Y	Y	743
GCA	GCT	GAA	ACA	GCT	GGA	AAT	GGA	TTA	TAT	AAC	TTT	ATT	GTT	CCT	CTC	AGG	GCA	TAT	TAT	2229
R	P	K	K	E	L	N	P	I	V	L	L	L	D	N	P	L	D	D	L	763
AGA	CCA	AAG	AAA	GAA	CTT	AAT	CCC	ATA	GTA	CTG	CTA	TTG	GAT	AAC	CCC	CTA	GAT	GAC	TTA	2289
L	R	C	G	V	T	F	A	A	N	M	V	V	V	D	K	E	S	T	M	783
CTC	AGG	TGT	GGA	GTG	ACT	TTT	GCT	GCT	AAT	ATG	GTG	GTT	GTG	GAT	AAA	GAG	AGC	ACC	ATG	2349
S	A	E	E	D	Y	M	A	D	A	K	T	I	V	N	V	Q	T	L	F	803
AGT	GCC	GAG	GAA	GAC	TAC	ATG	GCA	GAT	GCC	AAA	ACC	ATT	GTG	AAC	GTG	CAG	ACA	CTC	TTC	2409
R	L	F	S	S	L	S	I	I	T	E	L	T	H	P	A	N	M	R	F	823
AGG	TTG	TTT	TCC	AGT	CTC	AGT	ATT	ATC	ACA	GAG	CTA	ACT	CAC	CCC	GCC	AAC	ATG	AGA	TTC	2469
M	Q	F	R	A	K	D	C	Y	S	L	A	L	S	K	L	E	K	K	E	843
ATG	CAA	TTC	AGA	GCC	AAA	GAC	TGT	TAC	TCT	CTT	GCT	CTT	TCA	AAA	CTG	GAA	AAG	AAA	GAA	2529
R	E	R	G	S	N	L	A	F	M	F	R	L	P	F	A	A	G	R	V	863
CGG	GAG	AGA	GGC	TCT	AAC	TTG	GCC	TTT	ATG	TTT	CGA	CTG	CCT	TTT	GCT	GCT	GGG	AGG	GTG	2589
F	S	I	S	M	L	D	T	L	L	Y	Q	S	F	V	K	D	Y	M	I	883
TTT	AGC	ATC	AGT	ATG	TTG	GAC	ACT	CTG	CTG	TAT	CAG	TCA	TTT	GTG	AAG	GAT	TAT	ATG	ATT	2649
S	I	T	R	L	L	L	G	L	D	T	T	P	G	S	G	F	L	C	S	903
TCT	ATC	ACG	AGA	CTT	CTG	TTG	GGA	CTG	GAC	ACT	ACA	CCA	GGA	TCT	GGG	TTT	CTT	TGT	TCT	2709
M	K	I	T	A	D	D	L	W	I	R	T	Y	A	R	L	Y	Q	K	L	923
ATG	AAA	ATC	ACT	GCA	GAT	GAC	TTA	TGG	ATC	AGA	ACT	TAT	GCC	AGA	CTT	TAT	CAG	AAG	TTG	2769
C	S	S	T	G	D	V	P	I	G	I	Y	R	T	E	S	Q	K	L	T	943
TGT	TCT	TCT	ACT	GGA	GAT	GTT	CCC	ATT	GGA	ATC	TAC	AGG	ACT	GAG	TCT	CAG	AAA	CTT	ACT	2829
T	S	E	S	R	K	I	A	S	Q	S	Q	I	S	I	S	V	E	E	W	963
ACA	TCT	GAG	TCT	CGA	AAA	ATA	GCA	TCA	CAA	TCT	CAA	ATA	TCT	ATC	AGT	GTA	GAA	GAG	TGG	2889
E	D	T	K	D	S	K	E	Q	G	H	H	R	S	N	H	R	N	S	T	983
GAA	GAC	ACC	AAA	GAC	TCC	AAA	GAA	CAA	GGG	CAC	CAC	CGC	AGC	AAC	CAC	CGC	AAC	TCA	ACA	2949
S	S	D	Q	S	D	H	P	L	L	R	R	K	S	M	Q	W	A	R	R	1003
TCC	AGT	GAC	CAG	TCG	GAC	CAT	CCC	TTG	CTG	CGG	AGA	AAA	AGC	ATG	CAG	TGG	GCC	CGA	AGA	3009
L	S	R	K	G	P	K	H	S	G	K	T	A	E	K	I	T	Q	Q	R	1023
CTG	AGC	AGA	AAA	GGC	CCA	AAA	CAC	TCT	GGT	AAA	ACA	GCT	GAA	AAA	ATA	ACC	CAG	CAG	CGA	3069
L	N	L	Y	R	R	S	E	R	Q	E	L	A	E	L	V	K	N	R	M	1043
CTG	AAC	CTC	TAC	AGG	AGG	TCA	GAA	AGA	CAA	GAG	CTT	GCT	GAA	CTT	GTG	AAA	AAT	AGA	ATG	3129
K	H	L	G	L	S	T	V	G	Y	D	E	M	N	D	H	Q	S	T	L	1063
AAA	CAC	TTG	GGT	CTT	TCT	ACA	GTG	GGA	TAT	GAT	GAA	ATG	AAT	GAT	CAT	CAA	AGT	ACC	CTC	3189
S	Y	I	L	I	N	P	S	P	D	T	R	I	E	L	N	D	V	V	Y	1083
TCC	TAC	ATC	CTG	ATT	AAC	CCA	TCT	CCA	GAT	ACC	AGA	ATA	GAG	CTG	AAT	GAT	GTT	GTA	TAC	3249

FIGURE 12C

L	I	R	P	D	P	L	A	Y	L	P	N	S	E	P	S	R	R	N	S	1103
TTA	ATT	CGA	CCA	GAT	CCA	CTG	GCC	TAC	CTT	CCA	AAC	AGT	GAG	CCC	AGT	CGA	AGA	AAC	AGC	3309

I	C	N	V	T	G	Q	D	S	R	E	E	T	Q	L	*	1119
ATC	TGC	AAT	GTC	ACT	GGT	CAA	GAT	TCT	CGG	GAG	GAA	ACT	CAA	CTT	TGA	3357

TAAAAATAAAATGAGAAACTTTTTCTCTACAAAGACCTTGCTTGAAACCACAAAAGTTTTGCTGGCACGAAAGAACTA
GATGGAAATATATGTAATTCTCTCATATTTAAAAACGTAATCTCTTCTCTTAGAAGTATAGATCATTTTGAAACTTAAT
GTACTACTTACTGGTACTCTCCCTATTAATATTTGAAGGACCTCAATGGAATAAATTTGAAAAGCTAAATTTAAATACA
AAAATTTAAATCTGACATTTAATTGTTTTATAATAATCCAAACTCTATGAAAGCAATTTTAAAAATTATTAAGGTTTTA
TGAAGTTGACAAAATCTAACTATATTTGGTGCATCACAATGGACACAGAATGCTGCTGCTCCTCTTAAAAATTAAATGT
GTCATATTATATTCTTTAAACTTACTGTTTTACAAAATTGAGCTCATCGTAAATGTCTAGTCTTCTCACATAGAGATTA
ACCAACAAACTTGTGTGGCTGACTTTTGTGTAAGAATCATAGTTTGCTTTAGAATACAAATCTTTAAGTCATTTTAACT
TTTTTTCTGCCTTACGATATAAAAAATATTTATCTTAGAATTTGAGATGTTTCATAGCATGTTTTATTACATTGAAGAAA
CTAAAACATAAATGAAAAGAAACACTAGGTTCTGCACTTTTGGTAACTTTATGTCTAGCAAATATTTTATGCCAAGA
AAAGCATACTATAAAGCAAATATCTATTATTCTCCTAAACGAATGCCTAGCATAGAGAAAATACTTAATACACATTTGT
TGACTTAAATTTAATTCAAGGATTGAAAAATTAACGGATATCTTGAAATATACAGTAATGATTGTCCTTAGACTCTTG
AACTTTACCATCTTTCTATTTCATATATCTATATAGTAAATTTCACTAGAAAAATTCTTTTAAATTTGACAGAAGATAA
TTTATACCTTTTATGGACTCTGAAGACACTTCAAACATTAAAAAGTCCTTATGTCTTTGGTAATGAAACATACACTCAA
TGANGATGTATTAAATTTTGACTT

FIGURE 12D

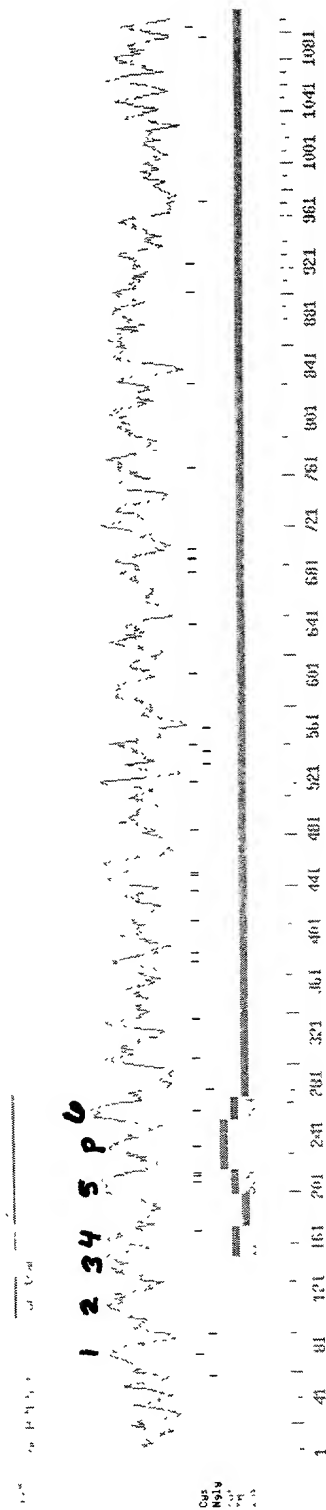


FIGURE 13

Protein Family / Domain Matches, HMMer version 2

Searching for complete domains in PFAM
hmmpfam - search a single seq against HMM database
HMMER 2.1.1 (Dec 1998)
Copyright (C) 1992-1998 Washington University School of Medicine
HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /prod/ddm/seqanal/PFAM/pfam6.4/Pfam
Sequence file: /prod/ddm/wspace/orfanal/oa-script.3743.seq

Query: Fbh54414

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
-----	-----	-----	-----	-----
ion_trans	Ion transport protein	62.4	9.9e-15	
1				

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
-----	-----	-----	-----	-----	-----	-----	-----
ion_trans	1/1	104	277	..	1 223 []	62.4	9.9e-15

Alignments of top-scoring domains:
ion_trans: domain 1 of 1, from 104 to 277: score 62.4, E = 9.9e-15

```
*->ilfildllfvllfllleivlkfiayglkstsniaakylksifnildll
      ++++ ++ + ++ +i+l + +y ++ ++ ++++i +il ++
Fbh54414 104 LWGLQVSVALISLFETILLGYLSYKGN-----IWEQILRIPFILEII 145

      ailplllllvflsgteqvakkrlrerfslelsqwyyrilrlLrllR
      ++p+++++++ ++ 1 + ++L ++
Fbh54414 146 NAVPFIISIFWPSLRN-----LFVPVFLNCW- 171

      lLrllrllrrletlfefelgtlaWslqslgralksilrflllllllligf
      l + + +++ + +l + ++r++ ++++ +l+l+ l++
Fbh54414 172 ---LAKHALENM-----INDL---HRAIQRTQSAMFNQVLILISTLLCL 209

      svigyllfkgyedlsenevdsnsefssyfdafyflfvltltvGfGdlvpv
      ++ + +++e+ ++ ++fd++yf++vt++tvGfGd++p+
Fbh54414 210 IFTCICIGIQLER-----IGKRLNLFDSLYFCIVTFSTVGFGDVTPE 251

      .wlgiiiffvlffiiivgllllnlliavi<-*
      +w++++f+ + +i+v+l+l + + +
Fbh54414 252 tWSSKLFV-VAMICVALVVLPIQFEQL 277
```

FIGURE 14

CLUSTAL W (1.74) multiple sequence alignment

```

54414.prot -----MVDLESEVPPLPPR---YRFRDL--LLGDEGWQN
AF089730 MARAKLPRSPSEKAGPGDTPAGSAAPEEPHGLSPLLPTRGGSVGSVDVQRLHVEDFSL
          : .. * * * * * *: * * * :
          TM1

54414.prot DD---RVQVEFYMNENTFKERLKLFFIKNQSSLRIFLNFSLKLLSCLLYIIRVLIENP
AF089730 DSSLSQVQVEFYVNENTFKERLKLFFIKNQSSLRIFLNFSLKLLTCLLYIVRVLIENP
          * : * * * * * : * * * * * : * * * * * : * * * * * : * * * * * :
          TM2

54414.prot SQGN-----EWSHIFWVNRSLFLWGLQVSVALISLFETILLGYLSY
AF089730 DQGIGCWGCTKYNYTFNGSSSEFHWAPILWVERKMA LWIQQVIVATISFLETMLLIYLSY
          ** : * * * * * : * * * * * : * * * * * : * * * * * :
          TM3 TM4

54414.prot KGNIWEQILRIPFILEIINAVPFIISIFWPSLRLNLFVFPVFLNCWLAKHALENMINDLHRA
AF089730 KGNIWEQIFHVSFVLEMINLTFIITVFWPPLRLNLFIPVFLNCWLAKHALENMINDFHRA
          * * * * * : * * * * * : * * * * * : * * * * * : * * * * * :
          TMS Pore

54414.prot IQRISAMFNQVLLILISTLLCLIFTICIGIHLERIGKRLNLFDSLYFCIVTFSTVGFSD
AF089730 ILRISAMFNQVLLILCTLLCLVFTGTGIGIHLERAGGNLNLTSFYFCIVTFSTVGFSD
          * * * * * : * * * * * : * * * * * : * * * * * : * * * * * :
          TM6

54414.prot VTETWSSKLFVVMAMICVALVVLPIQEQQLAYLWMERQKSGGNYSRHRAQTEKHVVLCVS
AF089730 VTEKIWPSQLLVVILICVTLVVLPIQFEELVYLWMERQKSGGNYSRHRARTEKHVVLCVS
          * * * * * : * * * * * : * * * * * : * * * * * : * * * * * :

54414.prot SLKIDLLMDFLNEFYAHPRLQDYVVVILCPTEMDVQVRRVLQIPMWSQRVIYLGQSALKD
AF089730 SLKIDLLMDFLNEFYAHPRLQDYVVVILCPSEMDVQVRRVLQIPLWSQRVIYLGQSALKD
          * * * * * : * * * * * : * * * * * : * * * * * : * * * * * :

54414.prot QDLIRAKMDDAEACFILSSRCEVDRTSSDHQTLIRAWAVKDFAPNCPLYVQILKPENKHF
AF089730 QDLIRAKMDNGEACFILSSRNEVDRTAADHQTILIRAWAVKDFAPNCPLYVQILKPENKHF
          * * * * * : * * * * * : * * * * * : * * * * * : * * * * * :

54414.prot IKFADHVVCEEFKYAMLALNCICPATSTLITLLVHTSRQEGQSQSPQWQRMYGRCSGN
AF089730 VKFADHVVCEECKYAMLALNCICPATSTLITLLVHTSRQEGQSQSPQWQRMYGRCSGN
          : * * * * * : * * * * * : * * * * * : * * * * * : * * * * * :

54414.prot EVYHIVLEESTFFAEYEGKSFTYASFHAKKFGVCLIGVRREDNKNILLNPGPRYIMNST
AF089730 EVYHIRMGDSKFFREYEGKSFTYAFAHAKKYGVCLIGLKREENKSIILLNPGPRHILAAS
          * * * * * : * * * * * : * * * * * : * * * * * : * * * * * :

54414.prot DICFYINITKEENSA--FKNQDQQRKSNVS--SFYHGPSRLPVHSIIASMGTVAILDQDT
AF089730 DTCFYINITKEENSAFIFKQEEKQNRRLAGQALYEGPSRLPVHSIIASM--VAMDLQNT
          * * * * * : * * * * * : * * * * * : * * * * * : * * * * * :

54414.prot SCRSASGPT-----LSLPTEGSKEIRRPISAPVLEVADTSSIQTCDLLSDQSEDETTP-
AF089730 DCRPSQGGSGGGGKLTLPTENGSGSRRPISAPVLELADSSALLPCDLLSDQSEDEVTPS
          . * * * * : * * * * * : * * * * * : * * * * * : * * * * * :

54414.prot DEEMSSNLEYAKGYPPSPYIGSSPTFCHLLHEKVFPFCLRLDKSCQHNYEDAKAYGFK
AF089730 DDEGLSVVEYKGYPPSPYIGSSPTLCHLLPVKAPFCCLRLDKGCKHNSYEDAKAYGFK
          * * * * * : * * * * * : * * * * * : * * * * * : * * * * * :

54414.prot NKLIIVAETAGNGLYNFIVPLRAYYRPKKELNPVLLLDN-P-----
AF089730 NKLIIISAETAGNGLYNFIVPLRAYYRSRRELNPVLLLDNPKDHHFLEAICCFPMVYYM
          * * * * * : * * * * * : * * * * * : * * * * * : * * * * * :

54414.prot -----LDDLRCGVTFANMNVVDKESTMSAEEDYMADAKTIVNVQTLFRLFSSLSIIT
AF089730 EGSVDNLDLSLLQCGIIYADNLVVVDKESTMSAEEDYMADAKTIVNVQTMFRLFPSLSIIT
          * * * * * : * * * * * : * * * * * : * * * * * : * * * * * :

```

G-hydrophobic-G
potassium channel
signature domain

FIGURE 15A

54414.prot ELTHPANMRMQFRAKDCYSLALSKEKKERERGSNLAFMFRLPFAAGRVSISMLDTLL
AF089730 ELTHPSNMRFMQFRAKDSYSLALSKEKQERENGSNLAFMFRLPFAAGRVSISMLDTLL
*****:*****:*****:***:*****:*****:*****

54414.prot YQSFVKDYMITITRLLGLDTPGSGFLCSMKITADDLWIRTYARLYQKLCSSSTGDVPIG
AF089730 YQSFVKDYMITITRLLGLDTPGSGYLCAMKVTFEDDLWIRTYGRLFQKLCSSSAEIPIG
*****:*****:***:***:*****:***:*****:***

54414.prot IYRTESQKLTTSERKIASQSQISISVEEWEDTKDSK-----EQGHRSNHRNSTSSD
AF089730 IYRTECH-VFSSEPHDLRAQSQISVNMEDCEDTREAKGPWGTRAASGGGTHGRHGGAD
*****.: : :***.: : :*****.:*: ***.:* . * : :*. **

54414.prot QSDHPLLRKSMQWARRLSRKGPKHSGK---TAEKITQQRNLNLYRRSERQFLAELVKNRM
AF089730 PVEHPLLRKSLQWARKLSRKSSKQAGKAPMTTDWITQQRSLYRRSERQELSELVKNRM
:*****:***:***:***:***:***:***:*****:*****

54414.prot KHLGLSTVGY-----DEMND-HQSTLSYILINPSPDTRIELNDV
AF089730 KHLGLPTTGVEDVANLTASDVMNRVNLGYLQDEMNDHHQNTLSYVLINPPDTRLEPNDI
*****.*.** ***** **.****:***:***:***:***:***:***

54414.prot VYLIRPDPLAYLPNSEPSRRNSICNVTG---QDSREETQL
AF089730 VYLIRSDPLAHTSSSQSRKSSCSNKLSSCNPETRDETQL
*****.***:***.* **.* . :*:***

FIGURE 15B

Input file Fbh53763pat.seq; Output File Fbh53763pat.tra
Sequence length 2847

```

CCACGCGTCCGGCCCTGTGCTTCGGATGGCGGCGGGAGGTTGATGGCGAGTGGTGCTGAAGGGACAGCTCCAGCAGTGG
CTGATTTGGGGGAGAAACAAAATCTGCAGATGGAATCCGAGCAGGGCGACTTCACCTTCAAGTGGTGAGCTCTCTCGAC
CTGCGGCCAGTCTCCACTCCATTCACGCCAGCCGATCTGCCCCTCCCGGAGGGGTCGGGCAGTGCCGGCTGGACCCG
CCCCGAGCTCCATGGTTTGCCCAACCCTGCGCGATGGTGACTCTGGGCGCGGAGGTTGGCGACTGGCAAATCCGCAGAT
CACAGAATGAAGGCGGGGAGCGCGCCGCGCGGGCGGGGGCTTTCTCCCCACCCAGCGCCAGGGAAGCGGCTCA
ACCACCTGAATCCGGAACACGCCAACAGTAGTTTCTCGTCGGAGAAGGGCGGCTCACCTGGGCGCCAAGACTCAGTCC
CGCTGCCCAGAGAACCTCGTCCACTCGGAAACCAAAGCAGAACCACCTTTCTCTCGGTCTCGTTAAGTCATGTCTGAGT

      M   G   K   I   E   N   N   E   R   V   I   L   N   V   G   G   T   R      18
CACAGAG ATG GGC AAG ATC GAG AAC AAC GAG AGG GTG ATC CTC AAT GTC GGG GGC ACC CGG      54
O
H   E   T   Y   R   S   T   L   K   T   L   P   G   T   R   L   A   L   L   A      38
CAC GAA ACC TAC CGC AGC ACC CTC AAG ACC CTG CCT GGA ACA CGC CTG GCC CTT CTT GCC      114
S   S   E   P   P   G   D   C   L   T   T   A   G   D   K   L   Q   P   S   P      58
TCC TCC GAG CCC CCA GGC GAC TGC TTG ACC ACG GCG GGC GAC AAG CTG CAG CCG TCG CCG      174
P   P   L   S   P   P   P   R   A   P   P   L   S   P   G   P   G   G   C   F      78
ECT CCA CTG TCG CCG CCG CCG AGA GCG CCC CCG CTG TCC CCC GGG CCA GGC GGC TGC TTC      234
E   G   G   A   G   N   C   S   S   R   G   G   R   A   S   D   H   P   G   G      98
GAG GGC GGC GCG GGC AAC TGC AGT TCC CGC GGC GGC AGG GCC AGC GAC CAT CCC GGT GGC      294
G   R   E   F   F   F   D   R   H   P   G   V   F   A   Y   V   L   N   Y   Y      118
GGC CGC GAG TTC TTC TTC GAC CGG CAC CCG GGC GTC TTC GCC TAT GTG CTC AAT TAC TAC      354
R   T   G   K   L   H   C   P   A   D   V   C   G   P   L   F   E   E   E   L      138
CGC ACC GGC AAG CTG CAC TGC CCC GCA GAC GTG TGC GGG CCG CTC TTC GAG GAG GAG CTG      414
A   F   W   G   I   D   E   T   D   V   E   P   C   C   W   M   T   Y   R   Q      158
GCC TTC TGG GGC ATC GAC GAG ACC GAC GTG GAG CCC TGC TGC TGG ATG ACC TAC CGG CAG      474
H   R   D   A   E   E   A   L   D   I   F   E   T   P   D   L   I   G   G   D      178
CAC CGC GAC GCC GAG GAG GCG CTG GAC ATC TTC GAG ACC CCC GAC CTC ATT GGC GGC GAC      534
P   G   D   D   E   D   L   A   A   K   R   L   G   I   E   D   A   A   G   L      198
CCC GGC GAC GAC GAG GAC CTG GCG GCC AAG AGG CTG GGC ATC GAG GAC GCG GCG GGG CTC      594
G   G   P   D   G   K   S   G   R   W   R   R   L   Q   P   R   M   W   A   L      218
GGG GGC CCC GAC GGC AAA TCT GGC CGC TGG AGG AGG CTG CAG CCC CGC ATG TGG GCC CTC      654
F   E   D   P   Y   S   S   R   A   A   R   F   I   A   F   A   S   L   F   F      238
TTC GAA GAC CCC TAC TCG TCC AGA GCC GCC AGG TTT ATT GCT TTT GCT TCT TTA TTC TTC      714
I   L   V   S   I   T   T   F   C   L   E   T   H   E   A   F   N   I   V   K      258
ATC CTG GTT TCA ATT ACA ACT TTT TGC CTG GAA ACA CAT GAA GCT TTC AAT ATT GTT AAA      774
N   K   T   E   P   V   I   N   G   T   S   V   V   L   Q   Y   E   I   E   T      278
AAC AAG ACA GAA CCA GTC ATC AAT GGC ACA AGT GTT GTT CTA CAG TAT GAA ATT GAA ACG      834

```

FIGURE 16A

D	P	A	L	T	Y	V	E	G	V	C	V	V	W	F	T	F	E	F	L	298
GAT	CCT	GCC	TTG	ACG	TAT	GTA	GAA	GGA	GTG	TGT	GTG	GTG	TGG	TTT	ACT	TTT	GAA	TTT	TTA	894
V	R	I	V	F	S	P	N	K	L	E	F	I	K	N	L	L	N	I	I	318
GTC	CGT	ATT	GTT	TTT	TCA	CCC	AAT	AAA	CTT	GAA	TTC	ATC	AAA	AAT	CTC	TTG	AAT	ATC	ATT	954
D	F	V	A	I	L	P	F	Y	L	E	V	G	L	S	G	L	S	S	K	338
GAC	TTT	GTG	GCC	ATC	CTA	CCT	TTC	TAC	TTA	GAG	GTG	GGA	CTC	AGT	GGG	CTG	TCA	TCC	AAA	1014
A	A	K	D	V	L	G	F	L	R	V	V	R	F	V	R	I	L	R	I	358
GCT	GCT	AAA	GAT	GTG	CTT	GGC	TTC	CTC	AGG	GTG	GTA	AGG	TTT	GTG	AGG	ATC	CTG	AGA	ATT	1074
F	K	L	T	R	H	F	V	G	L	R	V	L	G	H	T	L	R	A	S	378
TTC	AAG	CTC	ACC	CGC	CAT	TTT	GTA	GGT	CTG	AGG	GTG	CTT	GGA	CAT	ACT	CTT	CGA	GCT	AGT	1134
T	N	E	F	L	L	L	I	I	F	L	A	L	G	V	L	I	F	A	T	398
ACT	AAT	GAA	TTT	TTG	CTG	CTG	ATA	ATT	TTC	CTG	GCT	CTA	GGA	GTT	TTG	ATA	TTT	GCT	ACC	1194
M	I	Y	Y	A	E	R	V	G	A	Q	P	N	D	P	S	A	S	E	H	418
ATG	ATC	TAC	TAT	GCC	GAG	AGA	GTG	GGA	GCT	CAA	CCT	AAC	GAC	CCT	TCA	GCT	AGT	GAG	CAC	1254
T	Q	F	K	N	I	P	I	G	F	W	W	A	V	V	T	M	T	T	L	438
ACA	CAG	TTC	AAA	AAC	ATT	CCC	ATT	GGG	TTC	TGG	TGG	GCT	GTA	GTG	ACC	ATG	ACT	ACC	CTG	1314
G	Y	G	D	M	Y	P	Q	T	W	S	G	M	L	V	G	A	L	C	A	458
GGT	TAT	GGG	GAT	ATG	TAC	CCC	CAA	ACA	TGG	TCA	GGC	ATG	CTG	GTG	GGA	GCC	CTG	TGT	GCT	1374
L	A	G	V	L	T	I	A	M	P	V	P	V	I	V	N	N	F	G	M	478
CTG	GCT	GGA	GTG	CTG	ACA	ATA	GCC	ATG	CCA	GTG	CCT	GTC	ATT	GTC	AAT	AAT	TTT	GGA	ATG	1434
Y	Y	S	L	A	M	A	K	Q	K	L	P	R	K	R	K	K	H	I	P	498
TAC	TAC	TCC	TTG	GCA	ATG	GCA	AAG	CAG	AAA	CTT	CCA	AGG	AAA	AGA	AAG	AAG	CAC	ATC	CCT	1494
P	A	P	Q	A	S	S	P	T	F	C	K	T	E	L	N	M	A	C	N	518
CCT	GCT	CCT	CAG	GCA	AGC	TCA	CCT	ACT	TTT	TGC	AAG	ACA	GAA	TTA	AAT	ATG	GCC	TGC	AAT	1554
S	T	Q	S	D	T	C	L	G	K	D	N	R	L	L	E	H	N	R	S	538
AGT	ACA	CAG	AGT	GAC	ACA	TGT	CTG	GGC	AAA	GAC	AAT	CGA	CTT	CTG	GAA	CAT	AAC	AGA	TCA	1614
V	L	S	G	D	D	S	T	G	S	E	P	P	L	S	P	P	E	R	L	558
GTG	TTA	TCA	GGT	GAC	GAC	AGT	ACA	GGA	AGT	GAG	CCG	CCA	CTA	TCA	CCC	CCA	GAA	AGG	CTC	1674
P	I	R	R	S	S	T	R	D	K	N	R	R	G	E	T	C	F	L	L	578
CCC	ATC	AGA	CGC	TCT	AGT	ACC	AGA	GAC	AAA	AAC	AGA	AGA	GGG	GAA	ACA	TGT	TTC	CTA	CTG	1734
T	T	G	D	Y	T	C	A	S	D	G	G	I	R	K	G	Y	E	K	S	598
ACG	ACA	GGT	GAT	TAC	ACG	TGT	GCT	TCT	GAT	GGA	GGG	ATC	AGG	AAA	GGA	TAT	GAA	AAA	TCC	1794
R	S	L	N	N	I	A	G	L	A	G	N	A	L	R	L	S	P	V	T	618
CGA	AGC	TTA	AAC	AAC	ATA	GCG	GGC	TTG	GCA	GGC	AAT	GCT	CTG	AGG	CTC	TCT	CCA	GTA	ACA	1854
S	P	Y	N	S	P	C	P	L	R	R	S	R	S	P	I	P	S	I	L	638
TCA	CCC	TAC	AAC	TCT	CCT	TGT	CCT	CTG	AGG	CGC	TCT	CGA	TCT	CCC	ATC	CCA	TCT	ATC	TTG	1914
*																				639
TAA																				1917

FIGURE 16B

ACCAAACAACCAAACTGCATCAGTCGGCTAAATTGTATTAATTCAAGYGCTGTTTACCCCATTAATGGAAATAATTAAAT
GTAGAGTTACTCCAGGCTCCATTAATACAGTATAAATCTTGCGTGATACTACAATTTGAAGTCAGAAATGCCACTTGGG
TAGCTAATGAATCTTACCCAGGCTTTAAAGATTGTCTAAAGTAGTGCTAAGATCCCTCCTATTAATTGCCCTGATATCC
TTTTGCAATAAAATGACAGATAGTGTGTCAGATATTGACCAGTGCACTAATATATAAACATACCCTCAGGGAGATATATT
AAAACAGTGTGCTTCCAAATGCCAACCCTTCATTGGAACCTTATTTCTTGTGA

FOR "SECRET"

FIGURE 16C

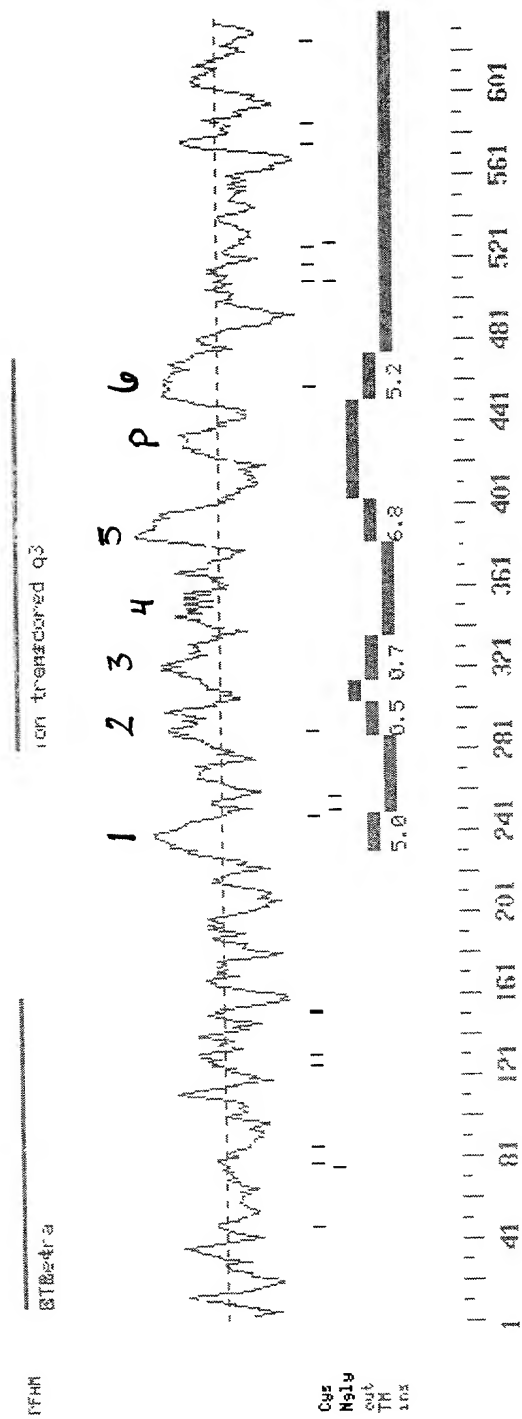


FIGURE 17

Protein Family / Domain Matches, HMMer version 2

Searching for complete domains in PFAM
hmmfam - search a single seq against HMM database
HMMER 2.1.1 (Dec 1998)
Copyright (C) 1992-1998 Washington University School of Medicine
HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /prod/ddm/seqanal/PFAM/pfam6.4/Pfam
Sequence file: /prod/ddm/wspace/orfanal/oa-script.4688.seq
Query: Fbh53763

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
K_tetra	K+ channel tetramerisation domain	156.7	4e-43	
ion_trans	Ion transport protein	116.9	3.9e-31	
oxidored_q3	NADH-ubiquinone/plastoquinone oxidoreduct	-81.7	5.6	

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
K_tetra	1/1	8	156	1	111	156.7	4e-43
oxidored_q3	1/1	317	467	1	177	-81.7	5.6
ion_trans	1/1	281	472	1	223	116.9	3.9e-31

Alignments of top-scoring domains:

K_tetra: domain 1 of 1, from 8 to 156: score 156.7, E = 4e-43

```

*->ErvrLNVGGkrFeTskstLtrfkpdTlLgrllktdsd.....
      Erv+LNVGG+r+eT++sTL ++ p T+L l++ s++++ ++ +
Fbh53763      8  ERVILNVGGTRHETYSRLKTL-PGTRLALLAS--SEppgdcclttag 51

.....vhearlrld
++ +++++ +++++ ++ +++++ ++ ++ +++++ d
Fbh53763      52 dklqpsppplsppprapplspppggcfeggagncssrggrA-----SD 94

fyddetgEyFFDRsPkhFetILnfYRtGdGkLhrp.evclidsfleEleFy
+ E+FFDR+P++F ++Ln+YRt GkLh+p +vc f+eEl+F+
Fbh53763      95 HPGGGR-EFFDRHPGVFAYVLNYYRT--GKLHCPaDVCGPLFEEELAFW 141

gldelaiesCcedeY<-*
g+de ++e+cc++Y
Fbh53763      142 GIDETDVEPCWMTY 156

```

oxidored_q3: domain 1 of 1, from 317 to 467: score -81.7, E = 5.6

```

*->mtiyivliLsillvlGflgVaskpsPiYgaLgLivaggvGCGlvsIG
+ +v iL + l +G+ g++sk + +++ +v ++ + + ++ l
Fbh53763      317 IIDFVAILPFYLEVGLSLSSKAAKDLVGLRVVRFVRI-LRIFKLT 362

gsFvalvlFLIYLGGMlVVfGytvalateeyPEaWgsnkvvtigdgval
Fv+l ++ g t t e+ + ++i ++l
Fbh53763      363 RHFVGLRVL-----GHTLRSTNEF-----L---LLI---IFL 389

vlgllievllvlgvl.....gwtvviivaltglGdwviYdvegsg
lg+li++ +++++ ++ +++++ e + +++ G w + v+ +
Fbh53763      390 ALGVLI FATMIYYAErvgaqndpSASEHTQFKNIP-IGFW--WAVVTM- 435

liredlsGvaaLYscgvvmfevaGwvLLvalfvvieltR<-*
++ G +Y +w + G L al +v+++++
Fbh53763      436 ----TTLGYGDMYPQ-TWSGMLVG--ALCALAGVLTIAM 467

```

FIGURE 18A

```

ion_trans: domain 1 of 1, from 281 to 472: score 116.9, E = 3.9e-31
      *->ilfildllfvllfllleivlkfiayglkstsniakylksifnildll
      l++++ ++v++f++e+++++ ++k      ++k+ ni+d+
Fbh53763 281  ALTYVEGVVCVWFTFEFLVRIVFSPNK-----LEFIKNLLNIIDFV 321

      ailplllllvflsgteqvakkrlrer.f.slelsq.wyyrilrflrlLr
      ailp++l ++l      +++++ ++ +      +flr++r
Fbh53763 322  AILPFYLEVGL-----SgLSKAAKdVL-----GFLRVVR 351

      llRlLrllrllrrletlfefelgtlaWslqslgralksilrfllllllll
      ++R  +lr++ +++      +++  l+ lg++l++ ++ +lll+++l
Fbh53763 352  FVR---ILRIFKLTR-----HFVG---LRVLGHTLRASTNEFLLLIIFL 389

      igfsvigyllfkgyedlse....nev dgnsefssyfdafyflfvltttvG
      +  +i++ + ++ e+      +++ +  +++++f +++ +f++++vt+tt+G
Fbh53763 390  ALGVLIFATMIYYAERVGAqndPSASEHTQFKNIPIGFWWAVVTMTTLG 439

      fGdlvpv.wlgiiffvlfiivgllllnllliavi<-*
      +Gd++p +w+g++++ ++++++g+l++++++vi
Fbh53763 440  YGDMYPQtWSGMLVG-ALCALAGVLTiAMPVPVI 472

//
Searching for complete domains in SMART
hmmpfam - search a single seq against HMM database
HMMER 2.1.1 (Dec 1998)
Copyright (C) 1992-1998 Washington University School of Medicine
HMMER is freely distributed under the GNU General Public License (GPL).
-----
HMM file: /ddm/robison/smart/smart/smart.all.hmms
Sequence file: /prod/ddm/wspace/orfanal/oa-script.4688.seq
-----
Query: Fbh53763

Scores for sequence family classification (score includes all domains):
Model Description Score E-value N
-----
BTB_4 72.7 7.8e-18 1

Parsed for domains:
Model Domain seq-f seq-t hmm-f hmm-t score E-value
-----
BTB_4 1/1 8 159 .. 1 114 [] 72.7 7.8e-18

Alignments of top-scoring domains:
BTB_4: domain 1 of 1, from 8 to 159: score 72.7, E = 7.8e-18
      *->cDvtlvvggdlggdnaegkkfhasqHkavLaacrrdSpyFkalfes.
      ++v+l+vvgg      ++ ++ +++L +      +++ +al+ s+
Fbh53763 8 ERVILNVGG-----TRHET--YRSTLCTL---PGTRLALLASs 40

      .....
      +++++ ++ +++ ++++++ +++++ ++ ++++++ +++ ++ +++++
Fbh53763 41 eppgdcclttagdklqpsppplsppprapplsppgggcfeggagncssrgg 90

      .....ieldDeallievspeaFralLnflyt.kldlpeedvenve
      + +++++++D      ++p +F +Ln+++t+kl+p + ++
Fbh53763 91 rasdhpgggrEFFFD-----RHPGVFAYVLNYYRTgKLHCPAD--VCGP 132

      elLelAdfldSYGqip.lvelCeefllknl<-*
      + e+ f+ G+ ++ ve C+++ +++
Fbh53763 133 LFEEELAFW---GIDeTdVEPCCWMTYRQH 159

//

```

FIGURE 18B

CLUSTAL W (1.74) multiple sequence alignment

Fbh53763pat
ratCIKE
MGKIENNERVILNVGGTRHETYRSTLKTLPGTRLALLASSEPPGDCLTTAGDKLQSPPPP
MGKIENNERVILNVGGTRHETYRSTLKTLPGTRLALLASSEPPQGDCLTAAGDKLQPLPPP
*****:*****

Fbh53763pat
ratCIKE
LSPPPRAPPLSPGPGGCFEGGAGNCSSRGGGRASDHPPGGGREFFFDHRHPGVFAYVLNYYRT
LSPPPRPPPLSPVPSGCFEGGAGNCSSHGGNGSDHPPGGGREFFFDHRHPGVFAYVLNYYRT
*****:*****

Fbh53763pat
ratCIKE
GKLHCPADVCGPLFEEELAFWGI DETDVEPCCWMTYRQHRDAEEALDIFETPD LIGGDPG
GKLHCPADVCGPLFEEELAFWGI DETDVEPCCWMTYRQHRDAEEALDIFETPD LIGGDPG

TM1

Fbh53763pat
ratCIKE
DDEDLAAKRLGIEDAAGLGGPDGKSGRWRRLQPRMWALFEDPYSSRAARFI AFASLFFIL
DDEDLGGKRLGIEDAAGLGGPDGKSGRWRRLQPRMWALFEDPYSSRAARFI AFASLFFIL
*****:*****
TM2

Fbh53763pat
ratCIKE
VSITTFCL⁺ETHEAFNIVKNKTEPVINGTSVVLQYEIETDPALTYVEGVCVVWFTFEFLVR
VSITTFCL⁺ETHEAFNIVKNKTEPVINGTSVVLQYEIETDPALTYVEGVCVVWFTFEFLVR
*****:*****
TM3 TM4

Fbh53763pat
ratCIKE
IVFSPNKLEFIKNLLNIIDFVAILPFYLEVGLSGLSSKAAKDVLGFLRVVRFVRIILRIFK
IVFSPNKLEFIKNLLNIIDFVAILPFYLEVGLSGLSSKAAKDVLGFLRVVRFVRIILRIFK
*****:*****
TM5 + + + +

Fbh53763pat
ratCIKE
LTRHFVGIRVLGHTLRASTNEFLLLIIFLALGVLI FATMIYYAERVGAQPNDPSASEHTQ
LTRHFVGIRVLGHTLRASTNEFLLLIIFLALGVLI FATMIYYAERVGAQPNDPSASEHTQ
*****:*****
+ Pore TM6

Fbh53763pat
ratCIKE
FKNIP⁺IGFWWAVVTMTTLGYGDMYPQTWSGMLVGALCALAGVLTIA MPVPVIVNNFGMY
FKNIP⁺IGFWWAVVTMTTLGYGDMYPQTWSGMLVGALCALAGVLTIA MPVPVIVNNFGMY
*****:*****

Fbh53763pat
ratCIKE
SLAMAKQKLPRKRKKHI PPAPQASSPTFCKTELNMACNSTQSDTCLGKDNRLLEHNRSVL
SLAMAKQKLPRKRKKHI PPAPLASSPTFCKTELNMACNSTQSDTCLGKENRLLEHNRSVL
*****:*****

Fbh53763pat
ratCIKE
SGDDSTGSEPPLSPPERLP⁺IRRSSTRDKNRRGETCFLLTTG DYTCASDGGIRKGYEKSRS
SGDDSTGSEPPLSPPERLP⁺IRRSSTRDKNRRGETCFLLTTG DYTCASDGGIRKGYEKSRS

Fbh53763pat
ratCIKE
LNNIAGLAGNALRLSPVTSPYN⁺SPCPLRRSRSPISIL
LNNIAGLAGNALRLSPVTSPYN⁺SPCPLRRSRSPISIL

FIGURE 19

Input file Fbh67076FL.seq; Output File Fbh67076FL.tra
Sequence length 6582

CCACGCGTCCGCCCCACGCGTCCGCCCCACGCGTCCGAGAAGGCTTAGGTGGGCAGGCAGGACGAGAGAAAGACTGAGAGG
AGGGAAAGCCCGCTAGGTGGGAGTACAGCGGCGCGAGGGTCGAGGGGGAACCTCGTCGGTGACAGATGAGGAGGGTGGG
CTTTCAGAACTAGTCCCCCTCGCACCCCGCCCCGCCCTCCCGCGCTGGGGTCTTCACGGTGCCCTGCCTCAGAGCCC
GGCTCCACCACGCCCCGAAGAGGGAGTCTGGCCGTCGGCTGGCTCAGGGCGGGCGGGTGGCTGTACCCAGGCTCCCTG
GCCCCAGTGCGGGACCAGAGCGCGGGGCGGGCGGGCAGCCGCGGGCCGAGGAGGGGCTGCGAGCGAAACGGCGCGGCGC
GGCACGGCGGACGAGTTAGGGCCGGGCGAGGGAGGCTGTGGCTCCCCGACAGAGACAGGGGAGTAGTGTGGGCTGAGG
M F R R S L N 7
CGAGACAGCCCGGTAGAGCCCAGCTCAGCGCCCGGCAGCCTTCGACGCG ATG TTC CGC CGG AGC TTG AAT 21
R F C A G E E K R V G T R T V F V G N H 27
CGT TTT TGT GCT GGA GAA GAG AAA CGA GTT GGC ACA CGC ACA GTG TTT GTT GGC AAT CAT 81
P V S E T E A Y I A Q R F C D N R I V S 47
CCA GTT TCG GAA ACA GAA GCT TAC ATT GCA CAA AGA TTT TGT GAT AAT AGA ATA GTC TCA 141
S K Y T L W N F L P K N L F E Q F R R I 67
TCT AAG TAT ACA CTT TGG AAT TTT CTC CCA AAG AAT CTG TTT GAA CAG TTT AGA AGA ATT 201
A N F Y F L I I F L V Q V T V D T P T S 87
GCA AAT TTT TAT TTT CTC ATA ATC TTC CTT GTA CAG GTC ACA GTA GAC ACA CCA ACT AGC 261
P V T S G L P L F F V I T V T A I K Q G 107
CCA GTT ACC AGT GGA CTT CCA CTT TTC TTT GTT ATA ACT GTT ACA GCC ATC AAG CAG GGA 321
Y E D W L R H R A D N E V N K S T V Y I 127
TAT GAG GAT TGG CTG AGA CAC AGA GCT GAC AAT GAA GTC AAC AAA AGC ACT GTT TAC ATT 381
I E N A K R V R K E S E K I K V G D V V 147
ATT GAA AAT GCA AAG CGA GTG AGA AAA GAA AGT GAA AAA ATC AAG GTT GGT GAT GTA GTA 441
E V Q A D E T F P C D L I L L S S C T T 167
GAA GTA CAG GCA GAT GAA ACC TTT CCC TGT GAT CTT ATT CTT CTA TCA TCT TGC ACC ACT 501
D G T C Y V T T A S L D G E S N C K T H 187
GAT GGA ACC TGT TAT GTC ACT ACA GCC AGT CTT GAT GGG GAA TCC AAT TGC AAG ACA CAT 561
Y A V R D T I A L C T A E S I D T L R A 207
TAT GCA GTA CGT GAT ACC ATT GCA CTG TGT ACA GCA GAA TCC ATC GAT ACC CTC CGA GCA 621
A I E C E Q P Q P D L Y K F V G R I N I 227
GCA ATT GAA TGT GAA CAG CCT CAA CCT GAC CTC TAC AAA TTT GTT GGG CGA ATC AAT ATC 681
Y S N S L E A V A R S L G P E N L L L K 247
TAC AGT AAT AGT CTT GAG GCT GTT GCC AGG TCT TTG GGA CCT GAA AAT CTC TTG CTG AAA 741
G A T L K N T E K I Y G V A V Y T G M E 267
GGA GCT ACG CTA AAA AAT ACC GAG AAG ATA TAT GGA GTT GCT GTT TAC ACT GGA ATG GAA 801

FIGURE 20A

T	K	M	A	L	N	Y	Q	G	K	S	Q	K	R	S	A	V	E	K	S	287
ACC	AAA	ATG	GCT	TTG	AAC	TAC	CAA	GGG	AAA	TCT	CAG	AAA	CGT	TCT	GCT	GTT	GAA	AAA	TCT	861
I	N	A	F	L	I	V	Y	L	F	I	L	L	T	K	A	A	V	C	T	307
ATT	AAT	GCT	TTC	CTG	ATT	GTA	TAT	TTA	TTT	ATC	TTA	CTG	ACC	AAA	GCT	GCA	GTA	TGC	ACT	921
T	L	K	Y	V	W	Q	S	T	P	Y	N	D	E	P	W	Y	N	Q	K	327
ACT	CTA	AAG	TAT	GTT	TGG	CAA	AGT	ACC	CCA	TAC	AAT	GAT	GAA	CCT	TGG	TAT	AAC	CAA	AAG	981
T	Q	K	E	R	E	T	L	K	V	L	K	M	F	T	D	F	L	S	F	347
ACT	CAG	AAA	GAG	CGA	GAG	ACC	TTG	AAG	GTT	TTA	AAA	ATG	TTC	ACC	GAC	TTC	CTA	TCA	TTT	1041
M	V	L	F	N	F	I	I	P	V	S	M	Y	V	T	V	E	M	Q	K	367
ATG	GTT	CTA	TTC	AAC	TTT	ATC	ATT	CCT	GTC	TCC	ATG	TAC	GTC	ACA	GTA	GAA	ATG	CAG	AAA	1101
F	L	G	S	F	F	I	S	W	D	K	D	F	Y	D	E	E	I	N	E	387
TTC	TTG	GGC	TCC	TTC	TTC	ATC	TCA	TGG	GAT	AAG	GAC	TTT	TAT	GAT	GAA	GAA	ATT	AAT	GAA	1161
G	A	L	V	N	T	S	D	L	N	E	E	L	G	Q	V	D	Y	V	F	407
GGA	GCC	CTG	GTT	AAC	ACA	TCA	GAC	CTT	AAT	GAA	GAA	CTT	GGT	CAG	GTG	GAT	TAT	GTA	TTT	1221
T	D	K	T	G	T	L	T	E	N	S	M	E	F	I	E	C	C	I	D	427
ACA	GAT	AAG	ACT	GGA	ACA	CTC	ACT	GAA	AAC	AGC	ATG	GAA	TTC	ATT	GAA	TGC	TGC	ATA	GAT	1281
G	H	K	Y	K	G	V	T	Q	E	V	D	G	L	S	Q	T	D	G	T	447
GGC	CAC	AAA	TAT	AAA	GGT	GTA	ACT	CAA	GAG	GTT	GAT	GGA	TTA	TCT	CAA	ACT	GAT	GGA	ACT	1341
L	T	Y	F	D	K	V	D	K	N	R	E	E	L	F	L	R	A	L	C	467
TTA	ACA	TAT	TTT	GAC	AAA	GTA	GAT	AAG	AAT	CGA	GAA	GAG	CTG	TTT	CTA	CGT	GCC	TTG	TGT	1401
L	C	H	T	V	E	I	K	T	N	D	A	V	D	G	A	T	E	S	A	487
TTA	TGT	CAT	ACT	GTA	GAA	ATC	AAA	ACA	AAC	GAT	GCT	GTT	GAT	GGA	GCT	ACA	GAA	TCA	GCT	1461
E	L	T	Y	I	S	S	S	P	D	E	I	A	L	V	K	G	A	K	R	507
GAA	TTA	ACC	TAT	ATC	TCC	TCT	TCA	CCA	GAT	GAA	ATA	GCT	TTG	GTG	AAA	GGA	GCT	AAA	AGG	1521
Y	G	F	T	F	L	G	N	R	N	G	Y	M	R	V	E	N	Q	R	K	527
TAC	GGG	TTC	ACA	TTT	TTA	GGA	AAT	CGA	AAT	GGA	TAT	ATG	AGA	GTA	GAG	AAC	CAA	AGA	AAA	1581
E	I	E	E	Y	E	L	L	H	T	L	N	F	D	A	V	R	R	R	M	547
GAA	ATA	GAA	GAA	TAT	GAA	CTT	CTT	CAC	ACC	TTA	AAC	TTT	GAT	GCT	GTC	CGG	CGA	CGT	ATG	1641
S	V	I	V	K	T	Q	E	G	D	I	L	L	F	C	K	G	A	D	S	567
AGT	GTA	ATT	GTG	AAG	ACT	CAA	GAA	GGA	GAC	ATA	CTT	CTC	TTT	TGT	AAA	GGA	GCA	GAC	TCG	1701
A	V	F	P	R	V	Q	N	H	E	I	E	L	T	K	V	H	V	E	R	587
GCA	GTT	TTT	CCC	AGA	GTG	CAA	AAT	CAT	GAA	ATT	GAG	TTA	ACT	AAA	GTC	CAT	GTG	GAA	CGT	1761
N	A	M	D	G	Y	R	T	L	C	V	A	F	K	E	I	A	P	D	D	607
AAT	GCA	ATG	GAT	GGG	TAT	CGG	ACA	CTC	TGT	GTA	GCC	TTC	AAA	GAA	ATT	GCT	CCA	GAT	GAT	1821
Y	E	R	I	N	R	Q	L	I	E	A	K	M	A	L	Q	D	R	E	E	627
TAT	GAA	AGA	ATT	AAC	AGA	CAG	CTC	ATA	GAG	GCA	AAA	ATG	GCC	TTA	CAA	GAC	AGA	GAA	GAA	1881
K	M	E	K	V	F	D	D	I	E	T	N	M	N	L	I	G	A	T	A	647
AAA	ATG	GAA	AAA	GTT	TTC	GAT	GAT	ATT	GAG	ACA	AAC	ATG	AAT	TTA	ATT	GGA	GCC	ACT	GCA	1941

FIGURE 20B

V	E	D	K	L	Q	D	Q	A	A	E	T	I	E	A	L	H	A	A	G	667
GTT	GAA	GAC	AAG	CTA	CAA	GAT	CAA	GCT	GCA	GAG	ACC	ATT	GAA	GCT	CTG	CAT	GCA	GCA	GGC	2001
L	K	V	W	V	L	T	G	D	K	M	E	T	A	K	S	T	C	Y	A	687
CTG	AAA	GTC	TGG	GTG	CTC	ACT	GGG	GAC	AAG	ATG	GAG	ACA	GCT	AAA	TCC	ACA	TGC	TAT	GCC	2061
C	R	L	F	Q	T	N	T	E	L	L	E	L	T	T	K	T	I	E	E	707
TGC	CGC	CTT	TTC	CAG	ACC	AAC	ACT	GAG	CTC	TTA	GAA	CTA	ACC	ACA	AAA	ACC	ATT	GAA	GAA	2121
S	E	R	K	E	D	R	L	H	E	L	L	I	E	Y	R	K	K	L	L	727
AGT	GAA	AGG	AAA	GAA	GAT	CGA	TTA	CAT	GAA	TTA	TTG	ATA	GAA	TAT	CGC	AAG	AAA	TTG	CTG	2181
H	E	F	P	K	S	T	R	S	F	K	K	A	W	T	E	H	Q	E	Y	747
CAT	GAG	TTT	CCT	AAA	AGT	ACT	AGA	AGC	TTT	AAA	AAA	GCA	TGG	ACA	GAA	CAT	CAG	GAA	TAT	2241
G	L	I	I	D	G	S	T	L	S	L	I	L	N	S	S	Q	D	S	S	767
GGA	TTA	ATC	ATA	GAT	GGC	TCC	ACA	TTG	TCA	CTC	ATA	CTA	AAT	TCT	AGT	CAA	GAC	TCT	AGT	2301
S	N	N	Y	K	S	I	F	L	Q	I	C	M	K	C	T	A	V	L	C	787
TCA	AAC	AAT	TAC	AAA	AGC	ATT	TTC	CTA	CAA	ATA	TGT	ATG	AAG	TGT	ACT	GCA	GTG	CTC	TGC	2361
C	R	M	A	P	L	Q	K	A	Q	I	V	R	M	V	K	N	L	K	G	807
TGT	CGG	ATG	GCA	CCA	TTA	CAG	AAA	GCC	CAG	ATT	GTC	AGA	ATG	GTG	AAG	AAT	TTA	AAA	GGC	2421
S	P	I	T	L	S	I	G	D	G	A	N	D	V	S	M	I	L	E	S	827
AGC	CCA	ATA	ACT	CTG	TCG	ATA	GGT	GAT	GGT	GCC	AAT	GAT	GTT	AGT	ATG	ATC	TTG	GAA	TCC	2481
H	V	G	I	G	I	K	G	K	E	G	R	Q	A	A	R	N	S	D	Y	847
CAT	GTG	GGA	ATA	GGT	ATT	AAA	GGC	AAA	GAA	GGT	CGC	CAA	GCA	GCT	AGG	AAT	AGC	GAT	TAT	2541
S	V	P	K	F	K	H	L	K	K	L	L	L	A	H	G	H	L	Y	Y	867
TCT	GTT	CCA	AAG	TTT	AAA	CAC	TTA	AAG	AAA	CTG	CTG	TTG	GCT	CAT	GGA	CAT	CTA	TAT	TAT	2601
V	R	I	A	H	L	V	Q	Y	F	F	Y	K	N	L	C	F	I	L	P	887
GTG	AGA	ATA	GCA	CAC	CTT	GTA	CAG	TAC	TTC	TTC	TAT	AAG	AAC	CTT	TGT	TTC	ATT	TTG	CCA	2661
Q	F	L	Y	Q	F	F	C	G	F	S	Q	Q	P	L	Y	D	A	A	Y	907
CAG	TTT	TTG	TAC	CAG	TTC	TTC	TGT	GGA	TTC	TCA	CAA	CAG	CCA	CTG	TAT	GAT	GCT	GCT	TAC	2721
L	T	M	Y	N	I	C	F	T	S	L	P	I	L	A	Y	S	L	L	E	927
CTT	ACA	ATG	TAC	AAT	ATC	TGC	TTC	ACA	TCC	TTG	CCC	ATC	CTG	GCC	TAT	AGT	CTA	CTG	GAA	2781
Q	H	I	N	I	D	T	L	T	S	D	P	R	L	Y	M	K	I	S	G	947
CAG	CAC	ATC	AAC	ATT	GAC	ACT	CTG	ACC	TCA	GAT	CCC	CGA	TTG	TAT	ATG	AAA	ATT	TCT	GGC	2841
N	A	M	L	Q	L	G	P	F	L	Y	W	T	F	L	A	A	F	E	G	967
AAT	GCC	ATG	CTA	CAG	TTG	GGC	CCC	TTC	TTA	TAT	TGG	ACA	TTT	CTG	GCT	GCC	TTT	GAA	GGG	2901
T	V	F	F	F	G	T	Y	F	L	F	Q	T	A	S	L	E	E	N	G	987
ACA	GTG	TTC	TTC	TTT	GGG	ACT	TAC	TTT	CTT	TTT	CAG	ACT	GCA	TCC	CTA	GAA	GAA	AAT	GGA	2961
K	V	Y	G	N	W	T	F	G	T	I	V	F	T	V	L	V	F	T	V	1007
AAG	GTA	TAC	GGA	AAC	TGG	ACT	TTT	GGA	ACC	ATT	GTT	TTT	ACA	GTC	TTA	GTA	TTC	ACT	GTA	3021
T	L	K	L	A	L	D	T	R	F	W	T	W	I	N	H	F	V	I	W	1027
ACC	CTG	AAG	CTT	GCC	TTG	GAT	ACC	CGA	TTC	TGG	ACG	TGG	ATA	AAT	CAC	TTT	GTG	ATT	TGG	3081

FIGURE 20C

G	S	L	A	F	Y	V	F	F	S	F	F	W	G	G	I	I	W	P	F	1047
GGT	TCT	TTA	GCC	TTC	TAT	GTA	TTT	TTC	TCA	TTC	TTC	TGG	GGA	GGA	ATT	ATT	TGG	CCT	TTT	3141
L	K	Q	Q	R	M	Y	F	V	F	A	Q	M	L	S	S	V	S	T	W	1067
CTC	AAG	CAA	CAG	AGA	ATG	TAT	TTT	GTA	TTT	GCC	CAA	ATG	CTG	TCT	TCT	GTA	TCC	ACA	TGG	3201
L	A	I	I	L	L	I	F	I	S	L	F	P	E	I	L	L	I	V	L	1087
TTG	GCT	ATA	ATT	CTT	CTA	ATA	TTT	ATC	AGC	CTG	TTC	CCT	GAG	ATT	CTT	CTG	ATA	GTA	TTA	3261
K	N	V	R	R	R	S	A	R	R	N	L	S	C	R	R	A	S	D	S	1107
AAG	AAT	GTA	AGA	AGA	AGA	AGT	GCC	AGG	AGA	AAT	CTG	AGC	TGT	AGA	AGG	GCA	TCT	GAC	TCA	3321
L	S	A	R	P	S	V	R	P	L	L	L	R	T	F	S	D	E	S	N	1127
TTA	TCC	GCC	AGA	CCT	TCA	GTC	AGA	CCT	CTT	CTT	TTA	CGA	ACA	TTC	TCA	GAC	GAA	TCT	AAT	3381
V	L	*																		1130
GTA	TTG	TAA																		3390

CAGAATCCGAATCTTGAACGCTATGTTATTGTCCTACAAGCATACTGACAGTGGTTACAGCTAAAAAGAAAGCATG
 AAGAAACAACTACAAAAAGTTATCATCTCAGGATACTTGATATGCAACACACTAAACCACCTCATGTCTAGAGTTAC
 AATAAATGTTTCATTAAATACCAAATGATTCTCTTAAGCATTTACCATTATTGTAAGTAGCCTTTATGGCCAAAGCTGT
 AAGTTAAGAATTATATGAAAGTTGAAAGCAAGAATACTTAGAATTCTGGCTTTAGTTAGAGTAATATAACTCAAATGGG
 TGCTCTTTTAACCCATGAACCTTGTGAATGGATTAAATACAATAGTATGAAGTAGAAGTTATGCAATGAGAATGAATA
 GATTTTGCTAATACTACTTTTTTGCCTGGCAGAAGAAATAGACTATTTGGATCACATTTCTCATTCCTCCTAAATGAT
 CATCTTAATTTTTTTTCCCAAGTACATAAGGAATACTTGAAAATACAGAATAACTAAATAGTATCAATGCATCAGACAG
 AATAGTTAATCCCTTCTGTTTACCCATGTGCTACTAATGTCTTGGTAGAATATTCTTGCCAAAAAATACCTTGAACGC
 TTATGTGGAAAGTGTTAACTTACGGGTATTTTTGTGGGAATAGAAAAAATGTTTATTTTTTATTCTTCTGAATTAAA
 CCCCACCTTATGGGTGTAAGCCTACTAGACTTGAAAATAAAGTATAAAACATTTCCAATCACTTAGTAGCCCCCTCAAAGT
 AGTTAGAAAAATAACAGATTTTTCCAGTGTTGATTTTACTGGGATCTGCAGTAAGGTGGTTTAAACCATAGTTATATAA
 AAATAAAGGTCATTCTGAATATCAGCCTTTTATAATTTTATGTGAAGAGGAAGAAATATAGCTTATTTTAACTTTTGA
 CGGTTTTTATTTGAAAGAGATTGCATTTATGCATATATGCAGTGCTTTTCTTAAACTTGGCCAATTTGGAAAGGGGGA
 AGGAGCCACCCCAAACGGTGGTTTCAGCTTGTTAGAGCCATGACTCTGTGAAGATGAATGTTGTCTCTTAACCTGGACAG
 GGAAATGGTCTAACTCTAAACCATGTAACGACCTTAGTAAAGTCCTTGACTAACTGAACTAGAAGGAAGGTTTAGCCT
 TCTAATTAGTTCACTTGAAACATAAATGTGAATGTCTTCAATGTTAAACACATACTTTTTTGGATATAAATGAC
 CATATTTATTTGACTGCTAGTTTTTTTGTTTTTTTTTTGTCTTTCTGGCATGCCTGTACTATTATTAATGTTTATATTG
 TACCTTGATTTGGAAAAGTATTGGAGTTAATCTGTATTATATTTATATAGTCCATATGGCACATTTGATTCTTCCACAT
 ATATTTTGTGTTAATGTTTAGGTATGATTTTTTTCTAAATCTAGAAAAGAACATAATTTCACTTATCAGAAGCCATTC
 CATCATTATAGACCCTTTTTCATTATTTTCATTTGCTCTCATATATCAGTATTATTTTGGAGCATTTTGTTACATGTCAT

FIGURE 20D

TCACAACTTACCTAAGTGTGCTGTGTTCTGGTAGCCCGTATTTGAGGTAAGCTGCTGAAAACAAAAGTCTCTATATTCT

TTGCCCTATTCCAAAGAGCTAAAAAGTCTAACCAGGAAAGCTTTTGATATTTTGTGTTTGTTCCTTGTCTTATGGT
TGTTGTTGCTGTATTATGATTGCTGTTTTACATAAAATCTATGGGAACGTGAATACAGACAAGAGAGCCACAGTAGAG
AGGCTTGTTTAATGCAGTACCATTGGAGAGTTAACAGAATAATCTAGTAGAAAAATAACTGGTTGCATGTAAATTCCT
TCCAGCCAGAAAGAAAGAAAGACAAGGAGTAAGGGGGATTTAGAGTTATGTCTCAGCTACACATTACATTGTGATACTG
CAGCTCAAATTCAGAATGGCAATGATACATGATATCATGGCCTAGATCCTTGAGAGGGACCTGGCTTTCCTTTTAAAA
GATATTTTACTGAAGAGCTAAAAACTGGCCAGTGTGGGGTTAGCAGATCGAATAAAGTGAATAGACCGTGCAGTATTC
CTAGCACTCAATGTAATCACCCCTATTTGTGACAGAGAAAGGGAAAAAATATAATAAGATCATCTACCTATAATTTGAA
TAATTTTGAGCTATCAAAATGTCTTTGTAATTTTCAACCGCTGTCCATTGTTTGAGGATGTTACCTACTAACTGAA
AACATTCATTCCATATCTACTTACACATACACCAGCAACAGTATAAATGTAAGCCTAACTTTGCAAAATTCGTAATAAT
TTAGTGATGGAATTTTAAATAACATGCAGTATATAAATGTGCAGATTTTATGCGTGTTGACAAAATCATTTCAGCT
TGCAAAATGGGACTGCAATATTACATTTTCACTTAAGCAGTTTTTTACATCTACGTTGTTGCTTTCCTAAATGAATGT
GAATGCCATCTTTTATGACTGCAACTTGCCTTTCCATTACAGAAATTTTGTGTTGATGTAATCAATAAACTTTGGTAT
GATAAA

FIGURE 20E

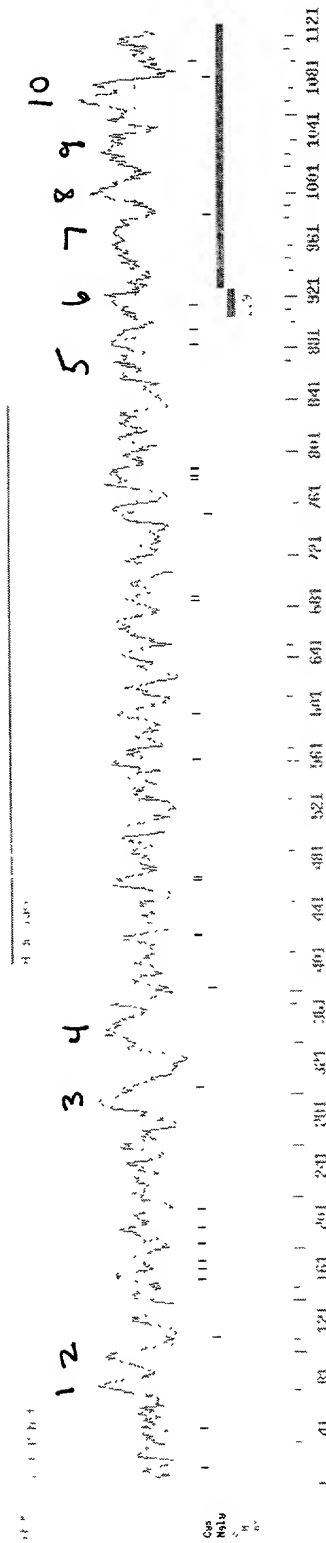


FIGURE 21

Protein Family / Domain Matches, HMMer version 2
 Searching for complete domains in PFAM
 hmmpfam - search a single seq against HMM database
 HMMER 2.1.1 (Dec 1998)
 Copyright (C) 1992-1998 Washington University School of Medicine
 HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /prod/ddm/seanal/PFAM/pfam6.4/Pfam
 Sequence file: /prod/ddm/wspace/orfanal/oa-script.13758.seq

Query: 67076

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
Hydrolase	haloacid dehalogenase-like hydrolase	12.7	0.019	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
Hydrolase	1/1	403	837	1	184	12.7	0.019

Alignments of top-scoring domains:

Hydrolase: domain 1 of 1, from 403 to 837: score 12.7, E = 0.019

```

      *->ikavvFDkDGTLTdkgkeppiaaivealrelgl.....
      ++ v+ Dk+GTLt+ + e +++ + +g++++ ++ ++++++
67076  403  VDYVFTDKTGTLTEN-SMEFIECCIDGHKYKGVtqevdglsgtdgtl 448

      ....apleeveklgrgl.gerilleggltaell.....
      + + + + e l +r+l + + + + t + + + + + + + + + +
67076  449  tyfdKVDKNREELFLRALcL--CHTVEIKTNDVdgesaeltyissp 496

      .....
      ++ ++ ++ + + + + + + + + + + + + + + + + + +
67076  497  deialvkgakrygftflgnrngymrvenqrkeieeyellhtlnfdavrrr 546

      .....
      + ++ +++ ++ ++ ++ ++ ++ ++ ++ ++ ++ ++ ++
67076  547  msvivktqegdillfckgadsavfprvqnheiltkvhvernamdgyrtl 596

      .....ld.evlgli
      ++ +++ ++ ++ + + + + + + + + + + + + + + + +
67076  597  cvafkeiapddyerinrqlieakmalqdreekmekvfdieTNmNLIGAT 646

      al.dklypgarealkaLkerGikvailTngdr.naealle....algla
      a++dkl + a++++aL+++G+kv++lT++ ++a+ + + + +
67076  647  AVeDKLQDQAAETIEALHAAGLKVWVLTGDKMeTAKSTCYacrIfQTNTE 696

      .lfdaivsdevggvgpvvvgKPkpeifllalerlgvkpeevg.....
      l + + + e + k + + + + l + + + + k + + + + +
67076  697  lLELTTKTIEESE-----RKEDRLHELLIEYRKLL--Hefpkstr 735

      .....
      + ++ + + + + + + + + + + + + + + + + + + + +
67076  736  sfkkawtehgeygliidgstlsliilnssqdsssnnyksiflqicmkctav 785

      .....p.kvlmvGDginDapalaaAGvgvamgn
      + + + + + + + + + + + + + + + + + + + + + + +
67076  786  lccrmaplqkaqivrmvknkgSpITLSIGDGANDVSMILESHVGIGIKG 835

      gg<-*

67076  836  KE 837
  
```

FIGURE 22

---MFRRLNRF CAGEEKRVGTRTVFVGN-HPVSETEAYIAQRFCDNRIVSSKYTLWNF
 MDCSLRLTLVRRYCAGEENWVDSRTIYVGHKEPPPGA EAYIPQRYPDNRIVSSKYTFWNF
 :.* :.*:*****: *.:*.:*.:* :.* :.*:*****:**.:*****:***

IPKKNLFEEQFRRIANFYFLIIFLVQVTVDIPTSPVTSGLPLFFVITVTAIKQGYEDWLRHR
IPKKNLFEEQFRRIANFYFLIIFLVQLIIDIPTSPVTSGLPLFFVITVTAIKQGYEDWLRHK
*****:*****:

ADNEVNKSTVYI I ENAKVRKESEKI KVGDVVEVQADETFPCDLILLSSCTTDGTCYVTI
ADNAMNQCPVHF IQHGKLVRKQSRKLRVGDI VMVKEDETFPCDLIFLSSNRADGTCHVTI

*** : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *

```
ASLDGESNCKTHYAVRDTIALCTAESIDTLRAAIECEQPQPDLYKFVGRINIYSNSLEAV
ASLDGESSHKTHYAVQDTKGFHTEADVDSLHATIECEQPQPDLYKFVGRINVYNDLNDPV
*****  *****.*  : *  :.*.*.*:*****:.*.:.*
```

ARSLGPENLLLKGATLKNTEKIYGVAVYTGME TKMALNYQGSQKRS AVEKSINAF L I V Y
VRPLGSENLLLRGATLKNTEKIFGVAIYTGME TKMALNYQSKSQKRS AVEKSMNTFL I V Y
* * * * * . * * * * * . * * * * * . * * * * * . * * * * * . * * * * *

LFILLTKAAVCTTLKYVWQSTPYNDEPWYNQKTQKERETLKVLMFTDFLSFMVLFNFI I
LCILVSKALINTVLKYVWQSEPFREDPWYNEKTESERQRNLFLLRAFTDFLAFMVLFNFI I
* * * * * . * * * * * * * . * * * * * . * * * * * . * * * * * . * * * * * . * * * * * . * * * * *

PVSMYVTTEMQKFLGSFFISWDKDFYDEEINEGALVNTSDLNEELGQVDVFIDKTGTLT
PVSMYVTTEMQKFLGSFYITWDEDMFDEEMGEPLVNTSDLNEELGQVEYIFTDKTGTLT
*****.

ENSMEFIECCIDGHKYKGVTDQVDGLSQTDTLTLYFDKVD---KNREELFLRALCLCHT
 ENNMAFKECCIEGHVYVPHVICNGQVLPDSSGIDMIDSSPGVCGREREELFFRAICLCHT
 ** * * ***** ** * . . . : . . . : * . . . : ***** ** : *****

VEIKTN---DAVDGATES---AELTYISSSPDEIALVKGAKRYGFTFLGNRNGYMRVENQ
VQVKDDHCGDDVDGPQKSPDAKSCVYISSSPDEVALVEGVQRLGFTYLRLKDNYPEILNR
:. : * *** :* *****:*:*:* :* ***:~* :** :*

RKEIEEYELLHTLNFDVRRRMSVIVKTQEGDILLFCKGADSAVFPRVQNHEIELTKVHV
ENDIERFELLEVLTFDSVRRRMSVIVKSTTGEIYLFCKGADSSIFPRVIEGKVDQVRSRV
: : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *

ERNAMDGVRTLCVAFKEIAPDDYERINRQLIEAKMALQDREEKMEKVFDDETNNMLIGA
ERNAVEGLRTLCAVAYKRLEPEQYEDACRLQLSAKVALQDREKKLAEAYEQIEKDLVLLGA
***** : ***** : * : * : * : * : * : * : * : *

D-SSSNYKSI FLQICMKCTAVLCCRMAPLQKAQIVRMVKNLKGSPITLSIGDGANDVSM
DGSSSGNYRELFLEICRNCSAVLCCRMAPLQKAQIVKLIKFSKEHPITLAI GDGANDVSM
* * * * * * * * * * * * * * * *

ILESHVGIGIKGKEGRQAARNSDYSVPKFKHLKKLLLAHGHLYYVRIAHLLVQYFFYKNLC

FIGURE 23A

'bh67076FL
louseAT1H

ILEAHVGIGVIGKEGRQAARNSDYAI PKFKHLKKMLLVHGHFYIRISELVQYFFYKNVC
 .**: *****.*****.***.***.***.***.*****.*

'bh67076FL
louseAT1H

FILPQFLYQF FCGFSQQPLYDAAYLTMYNICFTSLPI LAYSLL EQHINIDTLTSDPRLYM
FIFPQFLYQF FCGFSQQTLYDTAYLTLYNI SFTSLPI LLYSIMEQHVGIDVLKRDP TLYR
*.*****.*****.*****.*****.*****.*****.*****.

7bh67076FL
mouseAT1H

KISGNAMLQLGFLYWTFLLAAFE~~GTVFFFGTYFLE~~~~Q~~TASLEENGK~~VYG~~NWTFGTIVFTVL
DIAKNALLRWRVFIYWTFLLGVFDALVFFFGAYFI~~F~~ENTTVTINGQMFC~~NWTEGLVETVM~~

7bh67076FL
mouseAT1H

VFTVTILKLALDTRFWTWINHFVIWGSLAFYVFFSFFWGGIWPFLKQORMYFVFAQMLSS
VLTVTILKLALDTHYWWTWINHFVIWGSLLFYIAFSLLWGGVIWPFLSYQRMYYVFISMLSS
*.*****.:*****:***.*.....**

7bh67076FL
4ouseAT1H

VSTWLAIIILLIFISLFP EILLIVLK-----NVRRRSARRNLS CRRASDSL SAR
GPAWLGI ILLVTVGLLPDVLKKVLCRQLWPTATERTQNIQHQDSISEFTPLASLP SWGAQ
: * * * * * : * * : * * : * : * : * : * : * : *

-----PSVRPLLLRTFSDESNVL-----
 GSRLLAQCSSPSGRVVCsrWESEECpVLPLHPGLPHKARYGCCRSSLEMPt
 ** * : * * : **

FIGURE 23B

Input file Fbh67102FL.seq; Output File Fbh67102FL.tra
Sequence length 6074

CCACGCGTCCGGGAGGAGCGGAGGGAGAAGTAGGTTGCGAGCTCAGCACAGGCTCCGGCGCTGGCTCCCGCAGCTGAGT
TTGGGAGATGTCTAAGTGATTTTTTTTTTTTCCCGGAAGGCAAATGGCTGGCGTGGAAGCACAAACCCGCTTTCACTCTT
CGAATTTGTGCTTAGCTCTTTTCTTGTACCTTGCGACTCGTGACCAACATGCTGTGATGTGTGCCGAGGGAGGAATTGG

										M	T	E	A	L	Q	W	A	R	Y	10
TCAGCTACACAACCTGGATCTTACCACAGTTTGGAT	ATG	ACT	GAG	GCT	CTC	CAA	TGG	GCC	AGA	TAT										30
H	W	R	R	L	I	R	G	A	T	R	D	D	D	S	G	P	Y	N	Y	30
CAC	TGG	CGA	CGG	CTG	ATC	AGA	GGT	GCA	ACC	AGG	GAT	GAT	GAT	TCA	GGG	CCA	TAC	AAC	TAT	90
S	S	L	L	A	C	G	R	K	S	S	Q	I	P	K	L	S	G	R	H	50
TCC	TCG	TTG	CTC	GCC	TGT	GGG	CGC	AAG	TCC	TCT	CAG	ATC	CCT	AAA	CTG	TCA	GGA	AGG	CAC	150
R	I	V	V	P	H	I	Q	P	F	K	D	E	Y	E	K	F	S	G	A	70
CGG	ATT	GTT	GTT	CCC	CAC	ATC	CAG	CCC	TTC	AAG	GAT	GAG	TAT	GAG	AAG	TTC	TCC	GGA	GCC	210
Y	V	N	N	R	I	R	T	T	K	Y	T	L	L	N	F	V	P	R	N	90
TAT	GTG	AAC	AAT	CGA	ATA	CGA	ACA	ACA	AAG	TAC	ACA	CTT	CTG	AAT	TTT	GTG	CCA	AGA	AAT	270
L	F	E	Q	F	H	R	A	A	S	L	Y	F	L	F	L	V	V	L	N	110
TTA	TTT	GAA	CAA	TTT	CAC	AGA	GCT	GCC	AGT	TTA	TAT	TTC	CTG	TTC	CTA	GTT	GTC	CTG	AAC	330
W	V	P	L	V	E	A	F	Q	K	E	I	T	M	L	P	L	V	V	V	130
TGG	GTA	CCT	TTG	GTA	GAA	GCC	TTC	CAA	AAG	GAA	ATC	ACC	ATG	TTG	CCT	CTG	GTG	GTG	GTC	390
L	T	I	I	A	I	K	D	G	L	E	D	Y	R	K	Y	K	I	D	K	150
CTT	ACA	ATT	ATC	GCA	ATT	AAA	GAT	GGC	CTG	GAA	GAT	TAT	CGG	AAA	TAC	AAA	ATT	GAC	AAA	450
Q	I	N	N	L	I	T	K	V	Y	S	R	K	E	K	K	Y	I	D	R	170
CAG	ATC	AAT	AAT	TTA	ATA	ACT	AAA	GTT	TAT	AGT	AGG	AAA	GAG	AAA	AAA	TAC	ATT	GAC	CGA	510
C	W	K	D	V	T	V	G	D	F	I	R	L	S	C	N	E	V	I	P	190
TGC	TGG	AAA	GAC	GTT	ACT	GTT	GGG	GAC	TTT	ATT	CGC	CTC	TCC	TGC	AAT	GAG	GTC	ATC	CCT	570
A	D	M	V	L	L	F	S	T	D	P	D	G	I	C	H	I	E	T	S	210
GCA	GAC	ATG	GTA	CTA	CTC	TTT	TCC	ACT	GAT	CCA	GAT	GGA	ATC	TGT	CAC	ATT	GAG	ACT	TCT	630
G	L	D	G	E	S	N	L	K	Q	R	Q	V	V	R	G	Y	A	E	Q	230
GGT	CTT	GAT	GGA	GAG	AGC	AAT	TTA	AAA	CAG	AGG	CAG	GTG	GTT	CGG	GGA	TAT	GCA	GAA	CAG	690
D	S	E	V	D	P	E	K	F	S	S	R	I	E	C	E	S	P	N	N	250
GAC	TCT	GAA	GTT	GAT	CCT	GAG	AAG	TTT	TCC	AGT	AGG	ATA	GAA	TGT	GAA	AGC	CCA	AAC	AAT	750
D	L	S	R	F	R	G	F	L	E	H	S	N	K	E	R	V	G	L	S	270
GAC	CTC	AGC	AGA	TTC	CGA	GGC	TTC	CTA	GAA	CAT	TCC	AAC	AAA	GAA	CGC	GTG	GGT	CTC	AGT	810
K	E	N	L	L	L	R	G	C	T	I	R	N	T	E	A	V	V	G	I	290
AAA	GAA	AAT	TTG	TTG	CTT	AGA	GGA	TGC	ACC	ATT	AGA	AAC	ACA	GAG	GCT	GTT	GTG	GGC	ATT	870
V	V	Y	A	G	H	E	T	K	A	M	L	N	N	S	G	P	R	Y	K	310
GTG	GTT	TAT	GCA	GGC	CAT	GAA	ACC	AAA	GCA	ATG	CTG	AAC	AAC	AGT	GGG	CCA	CGG	TAT	AAG	930

FIGURE 24A

C	C	T	E	T	E	K	Q	H	G	D	A	G	L	L	N	G	K	A	E	710
TGC	TGC	ACA	GAA	ACA	GAG	AAA	CAA	CAC	GGT	GAT	GCA	GGC	CTC	CTG	AAT	GGC	AAG	GCA	GAG	2130
S	L	P	G	Q	P	L	A	C	N	L	C	Y	E	A	E	S	P	D	E	730
TCC	CTC	CCT	GGA	CAG	CCA	TTG	GCC	TGC	AAC	CTG	TGT	TAT	GAG	GCC	GAG	AGC	CCA	GAC	GAA	2190
A	A	L	V	Y	A	A	R	A	Y	Q	C	T	L	R	S	R	T	P	E	750
GCG	GCC	TTA	GTG	TAT	GCC	GCC	AGG	GCT	TAC	CAA	TGC	ACT	TTA	CGG	TCT	CGG	ACA	CCA	GAG	2250
Q	V	M	V	D	F	A	A	L	G	P	L	T	F	Q	L	L	H	I	L	770
CAG	GTC	ATG	GTG	GAC	TTT	GCT	GCT	TTG	GGA	CCA	TTA	ACA	TTT	CAA	CTC	CTA	CAC	ATC	CTG	2310
P	F	D	S	V	R	K	R	M	S	V	V	V	R	H	P	L	S	N	Q	790
CCC	TTT	GAC	TCA	GTA	AGA	AAA	AGA	ATG	TCT	GTT	GTG	GTC	CGA	CAC	CCT	CTT	TCC	AAT	CAA	2370
V	V	V	Y	T	K	G	A	D	S	V	I	M	E	L	L	S	V	A	S	810
GTT	GTG	GTG	TAT	ACG	AAA	GGC	GCT	GAT	TCT	GTG	ATC	ATG	GAG	TTA	CTG	TCG	GTG	GCT	TCC	2430
P	D	G	A	S	L	E	K	Q	Q	M	I	V	R	E	K	T	Q	K	H	830
CCA	GAT	GGA	GCA	AGT	CTG	GAG	AAA	CAA	CAG	ATG	ATA	GTA	AGG	GAG	AAA	ACC	CAG	AAG	CAC	2490
L	D	D	Y	A	K	Q	G	L	R	T	L	C	I	A	K	K	V	M	S	850
TTG	GAT	GAC	TAT	GCC	AAA	CAA	GGC	CTT	CGT	ACT	TTA	TGT	ATA	GCA	AAG	AAG	GTC	ATG	AGT	2550
D	T	E	Y	A	E	W	L	R	N	H	F	L	A	E	T	S	I	D	N	870
GAC	ACT	GAA	TAT	GCA	GAG	TGG	CTG	AGG	AAT	CAT	TTT	TTA	GCT	GAA	ACC	AGC	ATT	GAC	AAC	2610
R	E	E	L	L	L	E	S	A	M	R	L	E	N	K	L	T	L	L	G	890
AGG	GAA	GAA	TTA	CTA	CTT	GAA	TCT	GCC	ATG	AGG	TTG	GAG	AAC	AAA	CTT	ACA	TTA	CTT	GGT	2670
A	T	G	I	E	D	R	L	Q	E	G	V	P	E	S	I	E	A	L	H	910
GCT	ACT	GGC	ATT	GAA	GAC	CGT	CTG	CAG	GAG	GGA	GTC	CCT	GAA	TCT	ATA	GAA	GCT	CTT	CAC	2730
K	A	G	I	K	I	W	M	L	T	G	D	K	Q	E	T	A	V	N	I	930
AAA	GCG	GGC	ATC	AAG	ATC	TGG	ATG	CTG	ACA	GGG	GAC	AAG	CAG	GAG	ACA	GCT	GTC	AAC	ATA	2790
A	Y	A	C	K	L	L	E	P	D	D	K	L	F	I	L	N	T	Q	S	950
GCT	TAT	GCA	TGC	AAA	CTA	CTG	GAG	CCA	GAT	GAC	AAG	CTT	TTT	ATC	CTC	AAT	ACC	CAA	AGT	2850
K	D	A	C	G	M	L	M	S	T	I	L	K	E	L	Q	K	K	T	Q	970
AAA	GAT	GCC	TGT	GGG	ATG	CTG	ATG	AGC	ACA	ATT	TTG	AAA	GAA	CTT	CAG	AAG	AAA	ACT	CAA	2910
A	L	P	E	Q	V	S	L	S	E	D	L	L	Q	P	P	V	P	R	D	990
GCC	CTG	CCA	GAG	CAA	GTG	TCA	TTA	AGT	GAA	GAT	TTA	CTT	CAG	CCT	CCT	GTC	CCC	CGG	GAC	2970
S	G	L	R	A	G	L	I	I	T	G	K	T	L	E	F	A	L	Q	E	1010
TCA	GGG	TTA	CGA	GCT	GGA	CTC	ATT	ATC	ACT	GGG	AAG	ACC	CTG	GAG	TTT	GCC	CTG	CAA	GAA	3030
S	L	Q	K	Q	F	L	E	L	T	S	W	C	Q	A	V	V	C	C	R	1030
AGT	CTG	CAA	AAG	CAG	TTT	CTG	GAA	CTG	ACA	TCT	TGG	TGT	CAA	GCT	GTG	GTC	TGC	TGC	CGA	3090
A	T	P	L	Q	K	S	E	V	V	K	L	V	R	S	H	L	Q	V	M	1050
GCC	ACA	CCG	CTG	CAG	AAA	AGT	GAA	GTG	GTG	AAA	TTG	GTC	CGC	AGC	CAT	CTC	CAG	GTG	ATG	3150
T	L	A	I	G	D	G	A	N	D	V	S	M	I	Q	V	A	D	I	G	1070
ACC	CTT	GCT	ATT	GGT	GAT	GGT	GCC	AAT	GAT	GTT	AGC	ATG	ATA	CAA	GTG	GCA	GAC	ATT	GGG	3210

FIGURE 24C

I	G	V	S	G	Q	E	G	M	Q	A	V	M	A	S	D	F	A	V	S	1090
ATA	GGG	GTC	TCA	GGT	CAA	GAA	GGC	ATG	CAG	GCT	GTG	ATG	GCC	AGT	GAC	TTT	GCC	GTT	TCT	3270
Q	F	K	H	L	S	K	L	L	L	V	H	G	H	W	C	Y	T	R	L	1110
CAG	TTC	AAA	CAT	CTC	AGC	AAG	CTC	CTT	CTT	GTC	CAT	GGA	CAC	TGG	TGT	TAT	ACA	CGG	CTT	3330
S	N	M	I	L	Y	F	F	Y	K	N	V	A	Y	V	N	L	L	F	W	1130
TCC	AAC	ATG	ATT	CTC	TAT	TTT	TTC	TAT	AAG	AAT	GTG	GCC	TAT	GTG	AAC	CTC	CTT	TTC	TGG	3390
Y	Q	F	F	C	G	F	S	G	T	S	M	T	D	Y	W	V	L	I	F	1150
TAC	CAG	TTC	TTT	TGT	GGA	TTT	TCA	GGA	ACA	TCC	ATG	ACT	GAT	TAC	TGG	GTT	TTG	ATC	TTC	3450
F	N	L	L	F	T	S	A	P	P	V	I	Y	G	V	L	E	K	D	V	1170
TTC	AAC	CTC	CTC	TTC	ACA	TCT	GCC	CCT	CCT	GTC	ATT	TAT	GGT	GTT	TTG	GAG	AAA	GAT	GTG	3510
S	A	E	T	L	M	Q	L	P	E	L	Y	R	S	G	Q	K	S	E	A	1190
TCT	GCA	GAG	ACC	CTC	ATG	CAA	CTG	CCT	GAA	CTT	TAC	AGA	AGT	GGT	CAG	AAA	TCA	GAG	GCA	3570
Y	L	P	H	T	F	W	I	T	L	L	D	A	F	Y	Q	S	L	V	C	1210
TAC	TTA	CCC	CAT	ACC	TTC	TGG	ATC	ACC	TTA	TTG	GAT	GCT	TTT	TAT	CAA	AGC	CTG	GTC	TGC	3630
F	F	V	P	Y	F	T	Y	Q	G	S	D	T	D	I	F	A	F	G	N	1230
TTC	TTT	GTG	CCT	TAT	TTT	ACC	TAC	CAG	GGC	TCA	GAT	ACT	GAC	ATC	TTT	GCA	TTT	GGA	AAC	3690
P	L	N	T	A	A	L	F	I	V	L	L	H	L	V	I	E	S	K	S	1250
CCC	CTG	AAC	ACA	GCC	GCT	CTG	TTC	ATC	GTT	CTC	CTC	CAT	CTG	GTC	ATT	GAA	AGC	AAG	AGT	3750
L	T	W	I	H	L	L	V	I	I	G	S	I	L	S	Y	F	L	F	A	1270
TTG	ACT	TGG	ATT	CAC	TTG	CTG	GTC	ATC	ATT	GGT	AGC	ATC	TTG	TCT	TAT	TTT	TTA	TTT	GCC	3810
I	V	F	G	A	M	C	V	T	C	N	P	P	S	N	P	Y	W	I	M	1290
ATA	GTT	TTT	GGA	GCC	ATG	TGT	GTA	ACT	TGC	AAC	CCA	CCA	TCC	AAC	CCT	TAC	TGG	ATT	ATG	3870
Q	E	H	M	L	D	P	V	F	Y	L	V	C	I	L	T	T	S	I	A	1310
CAG	GAG	CAC	ATG	CTG	GAT	CCA	GTA	TTC	TAC	TTA	GTT	TGT	ATC	CTC	ACG	ACG	TCC	ATT	GCT	3930
L	L	P	R	F	V	Y	R	V	L	Q	G	S	L	F	P	S	P	I	L	1330
CTT	CTG	CCC	AGG	TTT	GTA	TAC	AGA	GTT	CTT	CAG	GGA	TCC	CTG	TTT	CCA	TCT	CCA	ATT	CTG	3990
R	A	K	H	F	D	R	L	T	P	E	E	R	T	K	A	L	K	K	W	1350
AGA	GCT	AAG	CAC	TTT	GAC	AGA	CTA	ACT	CCA	GAG	GAG	AGG	ACT	AAA	GCT	CTC	AAG	AAG	TGG	4050
R	G	A	G	K	M	N	Q	V	T	S	K	Y	A	N	Q	S	A	G	K	1370
AGA	GGG	GCT	GGA	AAG	ATG	AAT	CAA	GTG	ACA	TCA	AAG	TAT	GCT	AAC	CAA	TCA	GCT	GGC	AAG	4110
S	G	R	R	P	M	P	G	P	S	A	V	F	A	M	K	S	A	T	S	1390
TCA	GGA	AGA	AGA	CCC	ATG	CCT	GGC	CCT	TCT	GCT	GTA	TTT	GCA	ATG	AAG	TCA	GCA	ACT	TCC	4170
C	A	I	E	Q	G	N	L	S	L	C	E	T	A	L	D	Q	G	Y	S	1410
TGT	GCT	ATT	GAG	CAA	GGA	AAC	TTA	TCT	CTG	TGT	GAA	ACT	GCT	TTA	GAT	CAA	GGC	TAC	TCT	4230
E	T	K	A	F	E	M	A	G	P	S	K	G	K	E	S	*				1427
GAA	ACT	AAG	GCC	TTT	GAG	ATG	GCT	GGA	CCC	TCC	AAA	GGT	AAA	GAA	AGC	TAG				4281
ATACCCCTCCTTGAGTTGCAAGTATTCTTTCAAGGTTGGAAGAGGGATTTTGAAGAGGTATCTCTCCAAGCAAGAATGA																				
CTTGTTTTTCCATAAGGGACATGAGCATTTTACTAGGCTTGGAAGAGCTGACATGATGAGCATTATTGTATGTTTGTAT																				

FIGURE 24D

$\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \\ 13 \\ 14 \\ 15 \\ 16 \\ 17 \\ 18 \\ 19 \\ 20 \\ 21 \\ 22 \\ 23 \\ 24 \\ 25 \\ 26 \\ 27 \\ 28 \\ 29 \\ 30 \\ 31 \\ 32 \\ 33 \\ 34 \\ 35 \\ 36 \\ 37 \\ 38 \\ 39 \\ 40 \\ 41 \\ 42 \\ 43 \\ 44 \\ 45 \\ 46 \\ 47 \\ 48 \\ 49 \\ 50 \\ 51 \\ 52 \\ 53 \\ 54 \\ 55 \\ 56 \\ 57 \\ 58 \\ 59 \\ 60 \\ 61 \\ 62 \\ 63 \\ 64 \\ 65 \\ 66 \\ 67 \\ 68 \\ 69 \\ 70 \\ 71 \\ 72 \\ 73 \\ 74 \\ 75 \\ 76 \\ 77 \\ 78 \\ 79 \\ 80 \\ 81 \\ 82 \\ 83 \\ 84 \\ 85 \\ 86 \\ 87 \\ 88 \\ 89 \\ 90 \\ 91 \\ 92 \\ 93 \\ 94 \\ 95 \\ 96 \\ 97 \\ 98 \\ 99 \\ 100 \end{array}$

FIGURE 24E

TOPIC: CONCEPT

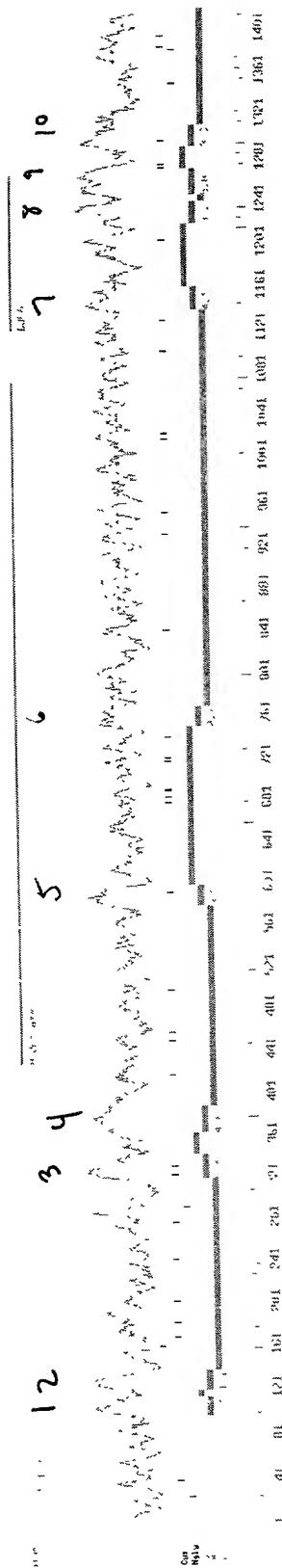


FIGURE 25

Protein Family / Domain Matches, HMMer version 2

Searching for complete domains in PFAM

hmmpfam - search a single seq against HMM database

HMMER 2.1.1 (Dec 1998)

Copyright (C) 1992-1998 Washington University School of Medicine

HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /prod/ddm/seqanal/PFAM/pfam6.4/Pfam

Sequence file: /prod/ddm/wspace/orfanal/oa-script.14482.seq

Query: 67102

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
Hydrolase	haloacid dehalogenase-like hydrolase	1.5	0.17	
1				
DUF6	Integral membrane protein DUF6	-24.6	9.4	
1				

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
Hydrolase	1/1	432	1077	1	184	1.5	0.17
DUF6	1/1	1127	1271	1	126	-24.6	9.4

Alignments of top-scoring domains:

Hydrolase: domain 1 of 1, from 432 to 1077: score 1.5, E = 0.17

```

*->ikavvFDkDGLtdgkeppiaaivealrelgl...apleevekll
    i  ++ Dk+GTLt+  + + ++ v+ ++ + ++ ++ + + e+
67102  432  IQYLFSDKTGTLTEN-KMVFRRCVAGFDYCHEenarRLESYQEAVS 477

    grgl.g..erilleggltaell.....
    + +   +++ ++l+ +++ ++ + ++ +++++ ++++++  +++ + +++
67102  478  EDEDfIdtVSGSLNMAKPRAPscrtvhnngplgnkpsnhlagssftlgsg 527

    .....
    ++ ++ +++++  +++ ++ +++++  ++ +  +++  + +++ ++
67102  528  egasevphsrqaafsspietdvvpdtrlldkfsqitprlfpmlpdtiqnp 577

    .....
    + ++  +  ++  + ++ ++ + +++++ ++ + ++ ++ ++
67102  578  pmetlyiidffialaicntvvvsapnqprqkirhpslgglpiksleeiks 627

    .....
    + + ++++++ ++++++ ++  ++ +  ++ ++ ++ +++ +
67102  628  lfqrwsrrsspslmsgkepssgvnafvsrlplfsrmkpaspvveevs 677

    .....
    +++  +++  +++++  +++ +  +++ ++ ++ +  +  ++
67102  678  qvcespqcssssacctetekqhgdagllngkaeslpqqplacnlcyeeas 727

    .....
    +++  +  + ++++++  +  ++ +  +  ++ ++
67102  728  pdeaalvyaarayqctlrsrtpeqvmvdfaalgpltfqllhilpfdsvrk 777

    .....
    + +  +++ ++  +++ ++  +  +  +++ + ++  +++
67102  778  rmsvvvrhplsnqvvytkgadsvimellsvaspdgaslekqqmivrekt 827

    .....
    ++ ++  + + ++  ++  +++++  +  +++  +++ +++++  +
67102  828  qkhlddyakqglrtlciakkvmsdteyaewlrnhflaetsidnreelle 877

```

FIGURE 26A

```

.....ld.evlglial.dklypgarealkaLkerGikvailTngdr.na
+ + ++ +lg+ +d l +g++e ++aL+++Gik++lT++ +++a
67102 878 samrleNKlTLLGATGieDRLQEGVPESIEALHKAGIKIWMLTGDKQeTA 927

ealle...alglal.lfdaivdsdevggvgpvvvgKPkpeifllalerlgv
+ ++ +l+ ++ + ++ + + + +g + i++++++ +
67102 928 VNIAYackLLEPDdKLFILNTQSKDA-----CGMLMSTILKELQKKTQA 971

kpeevg.....
pe+v+ +++ +++ +++++ + + +++++ + +++ +++ + ++
67102 972 LPEQVSlsedllqppvprdsgragliitgktlefalqeslqkqflelts 1021

.....p.kvlmvGDginDapalaaAGvgv
+ + ++ +++++ + +++ +l++GDg nD+ ++ A++g+
67102 1022 wcqavccratplqksevkvlrshlQvMTLAIGDGANDVSMIQVADIGI 1071

amnggg<-*

67102 1072 GVSGQE 1077

DUF6: domain 1 of 1, from 1127 to 1271: score -24.6, E = 9.4
*->fiWalytvfsskille..splttawrfliagilllilllflkkgppl
+++ +++++ +++ ++ + + ++ +l+ + ++ ++ + ++
67102 1127 LLFWYQFFCGFSGTSmtDYWVLIFNLLFTS--APPVIYGVLEKDV 1170

.....lallskilallylgilgtalgyllf
+ ++ + ++ ++++++ + ++ l +y +++++++y+ y
67102 1171 saetlmqlpelyrsgqkseaylpHTFWITL-LDAFYQSLVCFVPYFTYQ 1219

yalkyvsaskasvlsslsPvftlilsvllLgEkltlkqlGivlillGvl
+ ++ + +++++f++l ++ +lt++l i+ +l+ +
67102 1220 --GSDTDIFAfGNPLNTAALFIVLLHLVIESKSLTWIHLVLIIGSILSYF 1267

lisl<-*
l +
67102 1268 LFAI 1271

```

FIGURE 26B

[illegible]

FIGURE 27A

mouseAT5A KTONYLNLYAVEGLRTLCLIAKRVLSKEEYACWLQSHIEAEASVESREELLFQSAVRLETN
 ::: ** :**:::*. *** **::: **:::*****:::***:::

Fbh67102FL LTLGATGIEDRLQEGVPESIEALHKAGIKIWMLTGDKQETAVNIAYACKLLEPDDKLF
 mouseAT5A LHLLGATGIEDRLQEGVPETIAKLRQAGLQIWVLTGDKQETAINIAYACKLLDHGEEVIT
 * *****::: * *:***:::***:*****:*****: .:::

Fbh67102FL LNTQSKDACGMLMSTILKELQKKTQALPEQVS---LSEDLLQPPVPR--DSGLRAGLIIT
 mouseAT5A LNADSQEACAALLDQCLSYVQSRNPRSTLQNSESNLSVGFSFNPVSTSTDASPSPLVID
 :::*. *.. *. :*.. . * * ** .: ** *.. .*:

Fbh67102FL GKTLEFALQESLQKQFLELTSWCQAVVCCRATPLQKSEVVKLVRSHLQVMTLAIIGDGAND
 mouseAT5A GRSLAYALEKSLEDKFLFLAKQCRSVLCCRSTPLQKSMVVKLVRSKLKAMTLAIIGDGAND
 :: :***:::***:***. *.. *:::***:***** *****:*. *****

Fbh67102FL VSMIQVADIGIGVSGQEGMQAVMASDFAVSQFKHLSKLLLVHGHWCYTRLSNMILYFFYK
 mouseAT5A VSMIQVADVGVGISGQEGMQAVMASDFAVPRFRYLERLLIVHGHWCYSRLANMVLYFFYK
 *****:*. :*****:*****. *:***:*****:***:***:*****

Fbh67102FL NVAYVNLLFWYQFFCGFSGTSMTDYVWLIFFNLLFTSAPPVIYGVLEKDVSAETLMQLPE
 mouseAT5A NTMSVGLLFWFQFYCGFSASAMIDQWYLIFFNLLFSSLPQLVTGVLDKDV PADMLLREPQ
 * . *****:***:***.::* * * *****: * * : *::***:***: *:: *

Fbh67102FL LYRSGQKSEAYLPHTFWITLLDAFYQSLVCFVFPYFTYQGSDDTIFAFGNPLNTAALFIV
 mouseAT5A LYKSGQNMEEYRPRAFWLNLMVDAAFQSLVCFPIPYLAYYDSVDVETWGTPTVTAIALFTF
 ::*: * * *:***:***:***:*****:***: * .***:***:***:***: *

Fbh67102FL LLHLVIESKSLTWIHLVIIGSILSYFLFAIVFGAMCVTCNPPSPNPYWIMQEHMLDVFY
 mouseAT5A LLHLGIETKTWTWLNWLACGFSTFLFFSVALIYNTSCATCYPPSNPYWTMOTLLGDPLFY
 **** *::*: ***: * . * : * .::: * . * ***** * * : ***:

Fbh67102FL LVCILTTSIALLPFRVYRVLQGS LFPS-----PILR---AKHFDRLTPEERTK
 mouseAT5A LTCLIAPIAALLPRLFFKALQGS LFPTQLQLGRQLAKKPLNKFSDPKETFAQGQPPGHSE
 * .:::.. *****:***:*****: * : : * : * :

Fbh67102FL ALKKWR---G-----AGKMNQVTSKYAN--QSAGK-SGRRP-MPG-PSAVFA-MKS
 mouseAT5A TELSERKTMGPFETLPRDCASQASQFTQQLTCSPEASGEPSAVDTNMPLRENTLLEGLGS
 : . * * * .. * . * : : : * : * . * : : *

Fbh67102FL ATS-CAIEQGNLS-LCET-ALDQGYSETKAFEMAG---PSKGKES-----
 mouseAT5A QASGSSMPRGAISEVCPGDSKRQSSASQTARLSSLFHLPSFGSLNWISSLSLASGLGSV
 : * .::: * : * : * : * . * : : .::: * * * .

Fbh67102FL -----
 mouseAT5A LQLSGSSLQMDKQDGEFLSNPPQPEQDLHSFQGQVTGYL

FIGURE 27B

Input file Fbh44181pat.seq; Output File Fbh44181pat.tra
Sequence length 7221

GCCCGGGGATGGGAACGCGGCGCGGGGAGTGAGGCAGTGGCGGCGGCGGGTAAGCGGAAC TTCGCCCCGAGGGGCTC

GCCCGCTCCCGCCTCTGTCTTGTCTGGCCTCCACCTGCAGCCCCGCGGCCCCCGCGCCCCGCGGGACCCGGACGGCGACG

M W R W I R Q Q L G F D P P H Q S 17
ACGGGGGA ATG TGG CGC TGG ATC CGG CAG CAG CTG GGT TTT GAC CCA CCA CAT CAG AGT 51

D T R T I Y V A H R F P Q N G L Y T P Q 37
GAC ACA AGA ACC ATC TAC GTA GCC CAC AGG TTT CCT CAG AAT GGC CTT TAC ACA CCT CAG 111

K F I D N R I I S S K Y T V W N F V P K 57
AAA TTT ATA GAT AAC AGG ATC ATT TCA TCT AAG TAC ACT GTG TGG AAT TTT GTT CCA AAA 171

N L F E Q F R R V A N F Y F L I I F L V 77
AAT TTA TTT GAA CAG TTC AGA AGA GTG GCA AAC TTT TAT TTT CTT ATT ATA TTT TTG GTT 231

Q L M I D T P T S P V T S G L P L F F V 97
CAG CTT ATG ATT GAT ACA CCT ACC AGT CCA GTT ACC AGT GGA CTT CCA TTA TTC TTT GTG 291

I T V T A I K Q G Y E D W L R H N S D N 117
ATA ACA GTA ACT GCC ATA AAG CAG GGA TAT GAA GAT TGG TTA CGG CAT AAC TCA GAT AAT 351

E V N G A P V Y V V R S G G L V K T R S 137
GAA GTA AAT GGA GCT CCT GTT TAT GTT GTT CGA AGT GGT GGC CTT GTA AAA ACT AGA TCA 411

K N I R V G D I V R I A K D E I F P A D 157
AAA AAC ATT CGG GTG GGT GAT ATT GTT CGA ATA GCC AAA GAT GAA ATT TTT CCT GCA GAC 471

L V L L S S D R L D G S C H V T T A S L 177
TTG GTG CTT CTG TCC TCA GAT CGA CTG GAT GGT TCC TGT CAC GTT ACA ACT GCT AGT TTG 531

D G E T N L K T H V A V P E T A L L Q T 197
GAC GGA GAA ACT AAC CTG AAG ACA CAT GTG GCA GTT CCA GAA ACA GCA TTA TTA CAA ACA 591

V A N L D T L V A V I E C Q Q P E A D L 217
GTT GCC AAT TTG GAC ACT CTA GTA GCT GTA ATA GAA TGC CAG CAA CCA GAA GCA GAC TTA 651

Y R F M G R M I I T Q Q M E E I V R P L 237
TAC AGA TTC ATG GGA CGA ATG ATC ATA ACC CAA CAA ATG GAA GAA ATT GTA AGA CCT CTG 711

G P E S L L R G A R L K N T K E I F G 257
GGG CCG GAG AGT CTC CTG CTT CGT GGA GCC AGA TTA AAA AAC ACA AAA GAA ATT TTT GGT 771

V A V Y T G M E T K M A L N Y K S K S Q 277
GTT GCG GTA TAC ACT GGA ATG GAA ACT AAG ATG GCA TTA AAT TAC AAG AGC AAA TCA CAG 831

K R S A V E K S M N T F L I I Y L V I L 297
AAA CGA TCT GCA GTA GAA AAG TCA ATG AAT ACA TTT TTG ATA ATT TAT CTA GTA ATT CTT 891

I S E A V I S T I L K Y T W Q A E E K W 317
ATA TCT GAA GCT GTC ATC AGC ACT ATC TTG AAG TAT ACA TGG CAA GCT GAA GAA AAA TGG 951

D E P W Y N Q K T E H Q R N S S K I L R 337
GAT GAA CCT TGG TAT AAC CAA AAA ACA GAA CAT CAA AGA AAT AGC AGT AAG ATT CTG AGA 1011

FIGURE 28A

F	I	S	D	F	L	A	F	L	V	L	Y	N	F	I	I	P	I	S	L	357
TTT	ATT	TCA	GAC	TTC	CTT	GCT	TTT	TTG	GTT	CTC	TAC	AAT	TTC	ATC	ATT	CCA	ATT	TCA	TTA	1071
Y	V	T	V	E	M	Q	K	F	L	G	S	F	F	I	G	W	D	L	D	377
TAT	GTG	ACA	GTC	GAA	ATG	CAG	AAA	TTT	CTT	GGA	TCA	TTT	TTT	ATT	GGC	TGG	GAT	CTT	GAT	1131
L	Y	H	E	E	S	D	Q	K	A	Q	V	N	T	S	D	L	N	E	E	397
CTG	TAT	CAT	GAA	GAA	TCA	GAT	CAG	AAA	GCT	CAA	GTC	AAT	ACT	TCC	GAT	CTG	AAT	GAA	GAG	1191
L	G	Q	V	E	Y	V	F	T	D	K	T	G	T	L	T	E	N	E	M	417
CTT	GGA	CAG	GTA	GAG	TAC	GTG	TTT	ACA	GAT	AAA	ACT	GGT	ACA	CTG	ACA	GAA	AAT	GAG	ATG	1251
Q	F	R	E	C	S	I	N	G	M	K	Y	Q	E	I	N	G	R	L	V	437
CAG	TTT	CGG	GAA	TGT	TCA	ATT	AAT	GGC	ATG	AAA	TAC	CAA	GAA	ATT	AAT	GGT	AGA	CTT	GTA	1311
P	E	G	P	T	P	D	S	S	E	G	N	L	S	Y	L	S	S	L	S	457
CCC	GAA	GGA	CCA	ACA	CCA	GAC	TCT	TCA	GAA	GGA	AAC	TTA	TCT	TAT	CTT	AGT	AGT	TTA	TCC	1371
H	L	N	N	L	S	H	L	T	T	S	S	S	F	R	T	S	P	E	N	477
CAT	CTT	AAC	AAC	TTA	TCC	CAT	CTT	ACA	ACC	AGT	TCC	TCT	TTC	AGA	ACC	AGT	CCT	GAA	AAT	1431
E	T	E	L	I	K	E	H	D	L	F	F	K	A	V	S	L	C	H	T	497
GAA	ACT	GAA	CTA	ATT	AAA	GAA	CAT	GAT	CTC	TTC	TTT	AAA	GCA	GTC	AGT	CTC	TGT	CAC	ACT	1491
V	Q	I	S	N	V	Q	T	D	C	T	G	D	G	P	W	Q	S	N	L	517
GTA	CAG	ATT	AGC	AAT	GTT	CAA	ACT	GAC	TGC	ACT	GGT	GAT	GGT	CCC	TGG	CAA	TCC	AAC	CTG	1551
A	P	S	Q	L	E	Y	Y	A	S	S	P	D	E	K	A	L	V	E	A	537
GCA	CCA	TCG	CAG	TTG	GAG	TAC	TAT	GCA	TCT	TCA	CCA	GAT	GAA	AAG	GCT	CTA	GTA	GAA	GCT	1611
A	A	R	I	G	I	V	F	I	G	N	S	E	E	T	M	E	V	K	T	557
GCT	GCA	AGG	ATT	GGT	ATT	GTG	TTT	ATT	GGC	AAT	TCT	GAA	GAA	ACT	ATG	GAG	GTT	AAA	ACT	1671
L	G	K	L	E	R	Y	K	L	L	H	I	L	E	F	D	S	D	R	R	577
CTT	GGA	AAA	CTG	GAA	CGG	TAC	AAA	CTG	CTT	CAT	ATT	CTG	GAA	TTT	GAT	TCA	GAT	CGT	AGG	1731
R	M	S	V	I	V	Q	A	P	S	G	E	K	L	L	F	A	K	G	A	597
AGA	ATG	AGT	GTA	ATT	GTT	CAG	GCA	CCT	TCA	GGT	GAG	AAG	TTA	TTA	TTT	GCT	AAA	GGA	GCT	1791
E	S	S	I	L	P	K	C	I	G	G	E	I	E	K	T	R	I	H	V	617
GAG	TCA	TCA	ATT	CTC	CCT	AAA	TGT	ATA	GGT	GGA	GAA	ATA	GAA	AAA	ACC	AGA	ATT	CAT	GTA	1851
D	E	F	A	L	K	G	L	R	T	L	C	I	A	Y	R	K	F	T	S	637
GAT	GAA	TTT	GCT	TTG	AAA	GGG	CTA	AGA	ACT	CTG	TGT	ATA	GCA	TAT	AGA	AAA	TTT	ACA	TCA	1911
K	E	Y	E	E	I	D	K	R	I	F	E	A	R	T	A	L	Q	Q	R	657
AAA	GAG	TAT	GAG	GAA	ATA	GAT	AAA	CGC	ATA	TTT	GAA	GCC	AGG	ACT	GCC	TTG	CAG	CAG	CGG	1971
E	E	K	L	A	A	V	F	Q	F	I	E	K	D	L	I	L	L	G	A	677
GAA	GAG	AAA	TTG	GCA	GCT	GTT	TTC	CAG	TTC	ATA	GAG	AAA	GAC	CTG	ATA	TTA	CTT	GGA	GCC	2031
T	A	V	E	D	R	L	Q	D	K	V	R	E	T	I	E	A	L	R	M	697
ACA	GCA	GTA	GAA	GAC	AGA	CTA	CAA	GAT	AAA	GTT	CGA	GAA	ACT	ATT	GAA	GCA	TTG	AGA	ATG	2091
A	G	I	K	V	W	V	L	T	G	D	K	H	E	T	A	V	S	V	S	717
GCT	GGT	ATC	AAA	GTA	TGG	GTA	CTT	ACT	GGG	GAT	AAA	CAT	GAA	ACA	GCT	GTT	AGT	GTG	AGT	2151

FIGURE 28B

L	S	C	G	H	F	H	R	T	M	N	I	L	E	L	I	N	Q	K	S	737
TTA	TCA	TGT	GGC	CAT	TTT	CAT	AGA	ACC	ATG	AAC	ATC	CTT	GAA	CTT	ATA	AAC	CAG	AAA	TCA	2211
D	S	E	C	A	E	Q	L	R	Q	L	A	R	R	I	T	E	D	H	V	757
GAC	AGC	GAG	TGT	GCT	GAA	CAA	TTG	AGG	CAG	CTT	GCC	AGA	AGA	ATT	ACA	GAG	GAT	CAT	GTG	2271
I	Q	H	G	L	V	V	D	G	T	S	L	S	L	A	L	R	E	H	E	777
ATT	CAG	CAT	GGG	CTG	GTA	GTG	GAT	GGG	ACC	AGC	CTA	TCT	CTT	GCA	CTC	AGG	GAG	CAT	GAA	2331
K	L	F	M	E	V	C	R	N	C	S	A	V	L	C	C	R	M	A	P	797
AAA	CTA	TTT	ATG	GAA	GTT	TGC	AGA	AAT	TGT	TCA	GCT	GTA	TTA	TGC	TGT	CGT	ATG	GCT	CCA	2391
L	Q	K	A	K	V	I	R	L	I	K	I	S	P	E	K	P	I	T	L	817
CTG	CAG	AAA	GCA	AAA	GTA	ATA	AGA	CTA	ATA	AAA	ATA	TCA	CCT	GAG	AAA	CCT	ATA	ACA	TTG	2451
A	V	G	D	G	A	N	D	V	S	M	I	Q	E	A	H	V	G	I	G	837
GCT	GTT	GGT	GAT	GGT	GCT	AAT	GAC	GTA	AGC	ATG	ATA	CAA	GAA	GCC	CAT	GTT	GGC	ATA	GGA	2511
I	M	G	K	E	G	R	Q	A	A	R	N	S	D	Y	A	I	A	R	F	857
ATC	ATG	GGT	AAA	GAA	GGA	AGA	CAG	GCT	GCA	AGA	AAC	AGT	GAC	TAT	GCA	ATA	GCC	AGA	TTT	2571
K	F	L	S	K	L	L	F	V	H	G	H	F	Y	Y	I	R	I	A	T	877
AAG	TTC	CTC	TCC	AAA	TTG	CTT	TTT	GTT	CAT	GGT	CAT	TTT	TAT	TAT	ATT	AGA	ATA	GCT	ACC	2631
L	V	Q	Y	F	F	Y	K	N	V	C	F	I	T	P	Q	F	L	Y	Q	897
CTT	GTA	CAG	TAT	TTT	TTT	TAT	AAG	AAT	GTG	TGC	TTT	ATC	ACA	CCC	CAG	TTT	TTA	TAT	CAG	2691
F	Y	C	L	F	S	Q	Q	T	L	Y	D	S	V	Y	L	T	L	Y	N	917
TTC	TAC	TGT	TTG	TTT	TCT	CAG	CAA	ACA	TTG	TAT	GAC	AGC	GTG	TAC	CTG	ACT	TTA	TAC	AAT	2751
I	C	F	T	S	L	P	I	L	I	Y	S	L	L	E	Q	H	V	D	P	937
ATT	TGT	TTT	ACT	TCC	CTA	CCT	ATT	CTG	ATA	TAT	AGT	CTT	TTG	GAA	CAG	CAT	GTA	GAC	CCT	2811
H	V	L	Q	N	K	P	T	L	Y	R	D	I	S	K	N	R	L	L	S	957
CAT	GTG	TTA	CAA	AAT	AAG	CCC	ACC	CTT	TAT	CGA	GAC	ATT	AGT	AAA	AAC	CGC	CTC	TTA	AGT	2871
I	K	T	F	L	Y	W	T	I	L	G	F	S	H	A	F	I	F	F	F	977
ATT	AAA	ACA	TTT	CTT	TAT	TGG	ACC	ATC	CTG	GGC	TTC	AGT	CAT	GCC	TTT	ATT	TTC	TTT	TTT	2931
G	S	Y	L	L	I	G	K	D	T	S	L	L	G	N	G	Q	M	F	G	997
GGA	TCC	TAT	TTA	CTA	ATA	GGG	AAA	GAT	ACA	TCT	CTG	CTT	GGA	AAT	GGC	CAG	ATG	TTT	GGA	2991
N	W	T	F	G	T	L	V	F	T	V	M	V	I	T	V	T	V	K	M	1017
AAC	TGG	ACA	TTT	GGC	ACT	TTG	GTC	TTC	ACA	GTC	ATG	GTT	ATT	ACA	GTC	ACA	GTA	AAG	ATG	3051
A	L	E	T	H	F	W	T	W	I	N	H	L	V	T	W	G	S	I	I	1037
GCT	CTG	GAA	ACT	CAT	TTT	TGG	ACT	TGG	ATC	AAC	CAT	CTC	GTT	ACC	TGG	GGA	TCT	ATT	ATA	3111
F	Y	F	V	F	S	L	F	Y	G	G	I	L	W	P	F	L	G	S	Q	1057
TTT	TAT	TTT	GTA	TTT	TCC	TTG	TTT	TAT	GGA	GGG	ATT	CTC	TGG	CCA	TTT	TTG	GGC	TCC	CAG	3171
N	M	Y	F	V	F	I	Q	L	L	S	S	G	S	A	W	F	A	I	I	1077
AAT	ATG	TAT	TTT	GTG	TTT	ATT	CAG	CTC	CTG	TCA	AGT	GGT	TCT	GCT	TGG	TTT	GCC	ATA	ATC	3231
L	M	V	V	T	C	L	F	L	D	I	I	K	K	V	F	D	R	H	L	1097
CTC	ATG	GTT	GTT	ACA	TGT	CTA	TTT	CTT	GAT	ATC	ATA	AAG	AAG	GTC	TTT	GAC	CGA	CAC	CTC	3291

FIGURE 28C

H	P	T	S	T	E	K	A	Q	L	T	E	T	N	A	G	I	K	C	L	1117
CAC	CCT	ACA	AGT	ACT	GAA	AAG	GCA	CAG	CTT	ACT	GAA	ACA	AAT	GCA	GGT	ATC	AAG	TGC	TTG	3351
D	S	M	C	C	F	P	E	G	E	A	A	C	A	S	V	G	R	M	L	1137
GAC	TCC	ATG	TGC	TGT	TTC	CCG	GAA	GGA	GAA	GCA	GCG	TGT	GCA	TCT	GTT	GGA	AGA	ATG	CTG	3411
E	R	V	I	G	R	C	S	P	T	H	I	S	S	S	W	S	A	S	D	1157
GAA	CGA	GTT	ATA	GGA	AGA	TGT	AGT	CCA	ACC	CAC	ATC	AGC	AGT	TCA	TGG	AGT	GCA	TCG	GAT	3471
P	F	Y	T	N	D	R	S	I	L	T	L	S	T	M	D	S	S	T	C	1177
CCT	TTC	TAT	ACC	AAC	GAC	AGG	AGC	ATC	TTG	ACT	CTC	TCC	ACA	ATG	GAC	TCA	TCT	ACT	TGT	3531
*																				1178
TAA																				3534
AGGGGCAGTAGTACTTTGTGGGAGCCAGTTCACCTCCTTTTCTAAAATTCAGTGTGATCACCTGTTAATGGCCACACT																				
AGCTCTGAAATTAATTTCCAAAATCTTTGTAGTAGTTCATACCCACTCAGAGTTATAATGGCAAACAAACAGAAAGCAT																				
TAGTACAAGCCCCCTCCCAACACCCTTAATTTGAATCTGAACATGTTAAAATTTGAGAATAAAGAGACATTTTTCATCTC																				
TTTGTCTGGTTTGTCCCTTGTGCTTATGGGACTCCTAATGGCATTTCAGTCTGTTGCTGAGGCCATTATATTTTAATAT																				
AAATGTAGAAAAAGAGAGAAATCTTAGTAAAGAGTATTTTTTAGTATTAGCTTGATTATTGACTCTTCTATTTAAATC																				
TGCTTCTGTAAATTATGCTGAAAGTTTGCCCTGAGAACTCTATTTTTTTTATTAGAGTTATATTTAAAGCTTTTCATGGG																				
AAAAGTTAATGTGAATACTGAGGAATTTTGGTCCCTCAGTGACCTGTGTGTTAATTCATTAATGCATTCTGAGTTTAC																				
AGAGCAAATTAGGAGAATCATTTCCAACCATTTATTTACTGCAGTATGGGGAGTAAATTTATACCAATTCCTCTAACTGT																				
ACTGTAACACAGCCTGTAAAGTTAGCCATATAAATGCAAGGGTATATCATATATACAAATCAGGAATCAGGTCCGTTCA																				
CCGAACCTTCAAATTGATGTTTACTAATATTTTTGTGACAGAGTATAAAGACCCTATAGTGGGTAAATTAGATACTATTA																				
GCATATTATTAATTTAATGTCTTTATCATTTGGATCTTTTGCATGCTTTAATCTGGTTAACATATTTAAATTTGCTTTTTT																				
TTCTCTTTACCTGAAGGCTCTGTGTATAGTATTTTCATGACATCGTTGTACAGTTTAACTATCAATAAAAAGTTTGGACA																				
GTATTTAAATATTGCAAATATGTTTAATTATACAAATCAGAATAGTATGGGTAATTAAATGAATACAAAAGAAGAGCC																				
TCTTTCTGCAGCCGACTTAGACATGCTCTTCCCTTTCTATAAGCTAGATTTTAGAATAAAGGGTTTCAGTTAATAATCT																				
TATTTTCAGGTTATGTTCATCTAACTTATAGCAAACCTACCACAATACAGTGAGTTCTGCCAGTGTCCAGTACAAGGCAT																				
ATTTTCAGGTGTGGCTGTGGAATGTAAAAATGCTCAACTTGATCAGGTAATGTTAGCAATAAATTAAATGCTAAGAATG																				
ATTAATCGGGTACATGTTACTGTAATTAACCTATGCACTTCAAAACCTAACTTCCATCCTGAATTTATCAAGTAGTTC																				
AGTATTGTCATTTGTTTTTGTTTTATTGAAAAGTAATGTTGTCTTAAGATTTAGAAGTGATTATTAGCTTGAGAACTAT																				
TACCCAGCTCTAAGCAAATAATGATTGTATACATATTAAGATAATGGTTAAATGCGGTTTTTACCAAGTTTCCCTTGAA																				
AATGTAATTCCTTTTATGGAGATTTATTGTGCAGCCCTAAGCTTCCCTTCCATTTTCATGAATATAAGGCTTCTAGAATTG																				
GACTGGCAGGGGAAAGAATGGTAGAGACAGAAATTAAGACTTTATCCTTGTTTGTCTGTAAACTATTATTTCTTGCTA																				
ATGTAACATTTGTCTGTTCCAGTGATGTAAGGATATTAAGTTATTAAGCTAAATATTAATTTTCAAAAATAGTCCCTTCT																				

FIGURE 28D

[illegible]

FIGURE 28E

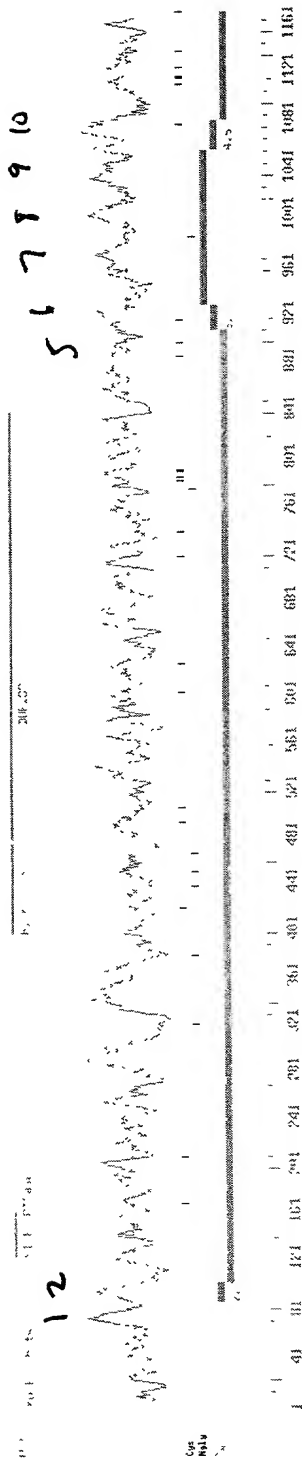


FIGURE 29

[illegible]

```

HMM file: /prod/ddm/seqanal/PFAM/pfam6.4/Pfam
Sequence file: /prod/ddm/wspace/orfanal/oa-script.15759.seq
-----
Query: 44181

```

Model	Description	Score	E-value	N
Hydrolase 1	haloacid dehalogenase-like hydrolase	42.8	8e-09	
E1-E2_ATPase 1	E1-E2 ATPase	8.6	0.13	
DUF132 1	Protein of unknown function DUF132	-72.9	9.4	

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
E1-E2 ATPase	1/1	126	164 ..	37	75 ..	8.6	0.13
DUF132	1/1	579	719 ..	1	160 []	-72.9	9.4
Hydrolase	1/1	401	842 ..	1	184 []	42.8	8e-09

E1-E2 ATPase: domain 1 of 1, from 126 to 164: score 8.6, E = 0.13

```
E1-E2_Alpase: domain 1 G1 I7 H10M L98V E166S  
                *->VLrDgKeeeiPaeeLvpGDiVevkpGdrVPADgrvvege<-*  
                V+r G++++ ++ +++GDIV+++ ++ PAD++++++  
    44181      126   VVRSGGLVKTRSKNIRVGDIVRIAKDEIFPADLVLLSSD       164
```

```

DUF132: domain 1 311, from 579 to 719: SC63_0127;
      *-->MeeliklVVIDTsVliaA.....LispkGlafkllelLfееKleN..
      M          V++ A++++L+ kG    +l    +++ +e ++
44181    579    MS-----VIVQapsgeklLLFAKGAESSILPKCIGGEIEKtr 614

```

```

      .YtSdeileEyifkillpKLecklpvEvslkkvl.vvlvskSkvinPRSF
      ++ + L+ + + l + +k+ + E      +++ + + +
44181  615 iHVDEFALKGL--RTLciAYRKfTSKE--YEEIDkRiFEARTALQQR--- 657

```

KESntkFncvCDpeDNKFLn...vvYesKAdvlITYdkDLLdRLRDNkK
 +k + F++++ + + A +L d+ R
 44181 658 ---EEKLA~~AAV~~----FQFIEkd~~l~~ILLGATA----VEDRLQDKVRETIEA 694

```

                                lkledHefkvLTPkEFiesveKkls<-*
                                l++ ++++ vLT                v +ls
44181    695 LRMAGIKVWVLTGDKHETAVSVSLS                    719

```

```
Hydrolase: domain 1 81 1, from 401 to 642: DQSLD LTRG; E  
          *->iakvvFDkDGTLtdgk.....  
          + v+ Dk+GTLT+ + + + + ++ + ++ +++ ++++++++  
44181    401    VEYVFDTKTGTLTENEmqfrecsingmkyqeingrlvpegptpsds 447
```

```

.....
++ + ++ ++ ++ ++ ++++++ ++++++++ +++++ + + ++
44181  448 gnl sylssslshlnnlshlttsssftrtspenete likeh dffkavslcht 497
.....
+ ++ +++ +++++ ++ + + + p ++a+vea++++g+ +

```

FIGURE 30A

'bh67076FL
louseAT1H
'bh44181

'bh67076FL
louseAT1H
'bh44181

'bh67076FL
louseATIH
'bh44181

bh67076FL
 mouseAT1H
 bh44181

7bh67076FL
4ouseAT1H
7bh44181

ARSLGPENLLLKGATLKNTEKIYGVAVYTGMETKMALNYQGSQKRSAVEKSINAH
VRPLGSENLLLRGATLKNTEKIFGVAIYTGME TKMALNYQS QSKRS AVEKSMNTH
VRPLGPESLLRGARLKNTKEIFGVAVYTGMETKMALNYKSKSQKRSAVEKSMNTH
* * * * *

Fbh67076FL
MouseAT1H
Fbh44181

LFILLTKAAVCTTLKYVWQSTPYNDEPWYNQKTQKERETLKVLMFTDFLSFMVLFNFII
LCILVSKALINTVLKYVWQSEPFREPWYNEKTESERQRNLFRAFTDFLAFMVLFNYII
LVILISEAVISTILKYTWQAEKWDEPWYNQKTEHQRNSSKILRFISDFLAFVLVLYNFII
* *

Fbh67076FL
MouseAT1H
Fbh44181

PVSMYVTVMQKFLGSFFISWDKDFYDEEINEGALVNTSDLNEELGQVDYVFIDDKTGTLT
PVSMYVTVMQKFLGSYFITWDEDMFDEEMGEGLVNTSDLNEELGQVEYIFIDDKTGTLT
PISLYVTVMQKFLGSFFIGWDLDLYHEESDQKAQVNTSDLNEELGQVEYVFIDDKTGTLT
*. . . * * * * *

Fbh67076FL
MouseAT1H
Fbh44181

ENSMEFIECCIDGHKYKG---VTQEV DGLSQT DGT LTYFDKVD-----
 ENNMAFKECCIEGHVYVPH--VICNGQVLPDSSG-IDMIDSSPGVC-----
 ENEMQFRECSINGMKYQEINGRLVPEGPTPDSSEGNLSYLSLSHLNLSHLTTSSSFRT
 ** * * * * . * * : : : * : : :

Fbh67076FL
MouseAT1H
Fbh44181

```

-----KNREELFLRALCLCHTVEIKTN---DAVDG---ATESAELTYISSSPDEIA
-----GREREELFFRAICLCHTVQVKDDHCGDDVDGPQK-SPDAKSCVYISSSPDEVA
SPENETELIKEHDLFFKA VSLCHTVQISNVQTDCTGDGPWQSNLAPSQLEY YASSPDEKA
      :.:**.:*:*****:..      **      . . . * :***** *

```

Fbh67076FL
MouseAT1H
Fbh44181

LVKGAKRYGFTFLGNRNGYMRVENQRKEIEEYELLHTLNFDVRRRMSVIVKTQEGDILL
LVEGVQRLGFTYLRDKNYMEILNRENDIERFELLEVLTFDSVRRRMSVIVKSTTGEIYL
LVEAAARIGIVFIGNSEETMEVKTG-KLERYKLLHILEFDSDRRRMSVIVQAPSGEKLL
*: : * : : : : * : : : * : * : : * : : * : : *

Fbh67076FL
MouseAT1H
Fbh44181

FCKGADSAVFPRVQNHEIELTKVHVERNAMDGYRTLCAVAFKEIAPDDYERINRQLIEAKM
FCKGADSSIFPRVIEGKVDQVRSRVERNAVEGLRTLCAVAYKRLEPEQYEDACRLLSAKV
FAKGAESSILPKCIGGEIEKTRIHVDEFALKGLRTLCIAYRKFTSKEYEEIDKRIFEART
*:****:*::*: ::::*: *::* *****:*.:::.:::** :::*:

Fbh67076FL
MouseAT1H

ALQDREEKMEKVDDIETNMNLIGATAVEDKLQDQAAETIEALHAAGLKVWVLTDGKMET
ALQDREKKLAEAYEOIEKDLVLLGATAVEDRLQEKAADTIEALQKAGIKVWVLTDGKMET

PS

Input file Fbh67084FL.seq; Output File Fbh67084FL.tra
Sequence length 4198

GGAGTCGACCCACGCGTCCGCATTGAGACAATGCCTCCACAAATACTTGATGCAAAATTCAGTAAGACAGCACTTGTTG

AATCACCATTATAGTTTCTGACAAATTGTTCTCAAAAAGGTACCAGCTGGAGGATGAGTCTGCGCATTGATGAA

M	P	L	M	M	S	E	E	G	F	E	N	E	E	S	D	Y	H	T	L	20
ATG	CCA	CTA	ATG	ATG	TCT	GAA	GAA	GGC	TTT	GAG	AAT	GAG	GAA	AGT	GAT	TAC	CAC	ACC	TTA	60
P	R	A	R	I	M	Q	R	K	R	G	L	E	W	F	V	C	D	G	W	40
CCA	CGA	GCC	AGG	ATA	ATG	CAA	AGG	AAA	AGA	GGA	CTG	GAG	TGG	TTT	GTC	TGT	GAT	GGC	TGG	120
K	F	L	C	T	S	C	C	G	W	L	I	N	I	C	R	R	K	K	E	60
AAG	TTC	CTC	TGT	ACC	AGT	TGC	TGT	GGT	TGG	CTG	ATA	AAT	ATT	TGT	CGA	AGA	AAG	AAA	GAG	180
L	K	A	R	T	V	W	L	G	C	P	E	K	C	E	E	K	H	P	R	80
CTG	AAA	GCT	CGC	ACA	GTA	TGG	CTT	GGA	TGT	CCT	GAA	AAG	TGT	GAA	GAA	AAA	CAT	CCC	AGG	240
N	S	I	K	N	Q	K	Y	N	V	F	T	F	I	P	G	V	L	Y	E	100
AAT	TCT	ATA	AAA	AAT	CAA	AAA	TAC	AAT	GTG	TTT	ACC	TTT	ATA	CCT	GGG	GTT	TTG	TAT	GAA	300
Q	F	K	F	F	L	N	L	Y	F	L	V	I	S	C	S	Q	F	V	P	120
CAA	TTC	AAG	TTT	TTT	TTG	AAT	CTC	TAT	TTT	CTA	GTG	ATA	TCC	TGC	TCA	CAG	TTT	GTA	CCA	360
A	L	K	I	G	Y	L	Y	T	Y	W	A	P	L	G	F	V	L	A	V	140
GCA	TTG	AAA	ATA	GGC	TAT	CTC	TAC	ACC	TAC	TGG	GCT	CCT	CTG	GGA	TTT	GTC	TTG	GCT	GTT	420
T	M	T	R	E	A	I	D	E	F	R	R	F	Q	R	D	K	E	V	N	160
ACT	ATG	ACA	CGG	GAA	GCA	ATT	GAT	GAA	TTT	CGG	CGT	TTT	CAG	CGT	GAC	AAG	GAA	GTG	AAT	480
S	Q	L	Y	S	K	L	T	V	R	G	K	V	Q	V	K	S	S	D	I	180
TCA	CAA	CTA	TAT	AGC	AAG	CTT	ACA	GTA	AGA	GGT	AAA	GTG	CAA	GTT	AAG	AGT	TCA	GAC	ATA	540
Q	V	G	D	L	I	I	V	E	K	N	Q	R	I	P	S	D	M	V	F	200
CAA	GTT	GGA	GAC	CTC	ATC	ATA	GTG	GAA	AAG	AAT	CAA	AGA	ATT	CCA	TCG	GAC	ATG	GTG	TTT	600
L	R	T	S	E	K	A	G	S	C	F	I	R	T	D	Q	L	D	G	E	220
CTT	AGG	ACT	TCA	GAA	AAA	GCA	GGT	TCG	TGT	TTT	ATT	CGA	ACT	GAT	CAA	CTA	GAT	GGT	GAA	660
T	D	W	K	L	K	V	A	V	S	C	T	Q	Q	L	P	A	L	G	D	240
ACT	GAC	TGG	AAG	CTG	AAG	GTG	GCA	GTG	AGC	TGC	ACG	CAA	CAG	CTG	CCG	GCT	CTG	GGG	GAC	720
L	F	S	I	S	A	Y	V	Y	A	Q	K	P	Q	M	D	I	H	S	F	260
CTT	TTT	TCT	ATC	AGT	GCT	TAT	GTT	TAT	GCT	CAG	AAA	CCA	CAA	ATG	GAC	ATT	CAC	AGT	TTC	780
E	G	T	F	T	R	E	D	S	D	P	P	I	H	E	S	L	S	I	E	280
GAA	GGC	ACA	TTT	ACC	AGG	GAA	GAC	AGT	GAC	CCG	CCC	ATT	CAT	GAA	AGT	CTC	AGC	ATA	GAA	840
N	T	L	W	A	S	T	I	V	A	S	G	T	V	I	G	V	V	I	Y	300
AAT	ACA	TTG	TGG	GCA	AGC	ACC	ATT	GTT	GCA	TCA	GGT	ACT	GTA	ATA	GGT	GTT	GTC	ATT	TAT	900
T	G	K	E	T	R	S	V	M	N	T	S	N	P	K	N	K	V	G	L	320
ACC	GGA	AAA	GAG	ACT	CGA	AGT	GTA	ATG	AAC	ACA	TCC	AAT	CCA	AAA	AAT	AAG	GTT	GGT	TTG	960
L	D	L	E	L	N	R	L	T	K	A	L	F	L	A	L	V	A	L	S	340
TTG	GAC	CTT	GAA	CTC	AAT	CGG	CTG	ACG	AAA	GCG	CTA	TTT	TTG	GCT	TTA	GTT	GCT	CTT	TCC	1020

FIGURE 32A

I	V	M	V	T	L	Q	G	F	V	G	P	W	Y	R	N	L	F	R	F	360
ATT	GTT	ATG	GTA	ACC	TTA	CAA	GGA	TTT	GTG	GGT	CCA	TGG	TAC	CGC	AAT	CTT	TTT	CGG	TTC	1080
L	L	L	F	S	Y	I	I	P	I	S	L	R	V	N	L	D	M	G	K	380
CTT	CTC	CTC	TTT	TCT	TAC	ATC	ATT	CCC	ATA	AGT	TTG	CGT	GTG	AAC	TTG	GAC	ATG	GGC	AAA	1140
A	V	Y	G	W	M	M	M	K	D	E	N	I	P	G	T	V	V	R	T	400
GCG	GTG	TAT	GGA	TGG	ATG	ATG	ATG	AAA	GAT	GAG	AAC	ATC	CCT	GGC	ACG	GTC	GTT	CGG	ACC	1200
S	T	I	P	E	E	L	G	R	L	V	Y	L	L	T	D	K	T	G	T	420
AGC	ACT	ATC	CCA	GAG	GAA	CTT	GGG	CGC	CTG	GTG	TAT	TTA	TTG	ACA	GAC	AAA	ACA	GGA	ACC	1260
L	T	Q	N	E	M	I	F	K	R	L	H	L	G	T	V	S	Y	G	A	440
CTC	ACC	CAG	AAT	GAA	ATG	ATA	TTT	AAG	CGG	CTG	CAC	CTG	GGC	ACC	GTG	TCC	TAT	GGC	GCC	1320
D	T	M	D	E	I	Q	S	H	V	R	D	S	Y	S	Q	M	Q	S	Q	460
GAC	ACG	ATG	GAT	GAG	ATC	CAG	AGC	CAT	GTC	AGG	GAC	TCC	TAC	TCA	CAG	ATG	CAG	TCT	CAA	1380
A	G	G	N	N	T	G	S	T	P	L	R	K	A	Q	S	S	A	P	K	480
GCT	GGT	GGA	AAC	AAT	ACT	GGT	TCA	ACT	CCA	CTA	AGA	AAA	GCC	CAA	TCT	TCA	GCT	CCC	AAA	1440
V	R	K	S	V	S	S	R	I	H	E	A	V	K	A	I	V	L	C	H	500
GTT	AGG	AAA	AGT	GTC	AGT	AGT	CGA	ATC	CAT	GAA	GCC	GTG	AAA	GCC	ATC	GTG	CTG	TGT	CAC	1500
N	V	T	P	V	Y	E	S	R	A	G	V	T	E	E	T	E	F	A	E	520
AAC	GTG	ACC	CCC	GTG	TAT	GAG	TCT	CGG	GCC	GGC	GTT	ACT	GAG	GAG	ACT	GAG	TTC	GCA	GAG	1560
A	D	Q	D	F	S	D	E	N	R	T	Y	Q	A	S	S	P	D	E	V	540
GCT	GAC	CAA	GAC	TTC	AGT	GAT	GAG	AAT	CGC	ACC	TAC	CAG	GCT	TCC	AGC	CCG	GAT	GAG	GTC	1620
A	L	V	Q	W	T	E	S	V	G	L	T	L	V	S	R	D	L	T	S	560
GCT	CTG	GTG	CAG	TGG	ACA	GAG	AGT	GTG	GGC	CTC	ACG	CTG	GTC	AGC	AGG	GAC	CTC	ACC	TCC	1680
M	Q	L	K	T	P	S	G	Q	V	L	S	F	C	I	L	Q	L	F	P	580
ATG	CAG	CTG	AAG	ACC	CCC	AGT	GGC	CAG	GTC	CTC	AGC	TTC	TGC	ATT	CTG	CAG	CTG	TTT	CCC	1740
F	T	S	E	S	K	R	M	G	V	I	V	R	D	E	S	T	A	E	I	600
TTC	ACC	TCC	GAG	AGC	AAG	CGG	ATG	GGC	GTC	ATC	GTC	AGG	GAT	GAA	TCC	ACG	GCA	GAA	ATC	1800
T	F	Y	M	K	G	A	D	V	A	M	S	P	I	V	Q	Y	N	D	W	620
ACA	TTC	TAC	ATG	AAG	GGC	GCT	GAC	GTG	GCC	ATG	TCT	CCT	ATC	GTG	CAG	TAT	AAT	GAC	TGG	1860
L	E	E	E	C	G	N	M	A	R	E	G	L	R	T	L	V	V	A	K	640
CTG	GAA	GAG	GAG	TGC	GGA	AAC	ATG	GCT	CGC	GAA	GGA	CTG	CGG	ACC	CTC	GTG	GTT	GCA	AAG	1920
K	A	L	T	E	E	Q	Y	Q	D	F	E	S	R	Y	T	Q	A	K	L	660
AAG	GCG	TTG	ACA	GAG	GAG	CAG	TAC	CAG	GAC	TTT	GAG	AGC	CGA	TAC	ACT	CAA	GCC	AAG	CTG	1980
S	M	H	D	R	S	L	K	V	A	A	V	V	E	S	L	E	R	E	M	680
AGC	ATG	CAC	GAC	AGG	TCC	CTC	AAG	GTG	GCC	GCG	GTA	GTC	GAG	AGC	CTG	GAG	AGG	GAG	ATG	2040
E	L	L	C	L	T	G	V	E	D	Q	L	Q	A	D	V	R	P	T	L	700
GAA	CTG	CTG	TGC	CTC	ACC	GGC	GTG	GAG	GAC	CAG	CTG	CAG	GCA	GAC	GTG	CGG	CCC	ACG	CTG	2100
E	M	L	R	N	A	G	I	K	I	W	M	L	T	G	D	K	L	E	T	720
GAG	ATG	CTG	CGC	AAC	GCC	GGG	ATC	AAG	ATA	TGG	ATG	CTA	ACA	GGC	GAT	AAA	CTC	GAG	ACA	2160

FIGURE 32B

102453 129
T02T2T "E3420T

GGTTTGCCATTGCTACCAAGCAAGCACCACAAGAAAGGGAGGGTACGCCAGGCGAGCCCAGGGCACAGATGCTGAGACA
GCCTCTCCTTCTCAGTGCAGGGACGTACCCCTGCCAGGCAAGCCCAGGGCACAGATGCCAGGATGGCTTCTCCCTCTC
AGTGCGAGGCTTCACCCCTGCCAGGCAAGCCCAGGGCATAGATGCTGAGACAGCCTCTCCCTCTCAGTGCAGGGACGTC
ACCCCTGCCAGGCAAGCCCAGGGCACAGAGGCGGGACGGCCTCTCCCTCTCAGTGTGAGGCTTCACCCATGCTAGGCA
AGCCCAGGGCACAGATGCCGGGATGGCCCCCTCCCTCTCAGTGCGGGAACGTACCCCTGCCAGGCAAGCCCAGGGCACA
GATGCTGCGATGGCCTCTTCTCTTAAGTGTGGGGCCTCACCCCTGCTTTTCTTTCTTTTTTTGTATTGTCAAAATTGT
ATTTCCATATTGAAGCAGCTTGAGTTTCTACTGAAAATGAGCCCGAATTATTTCACTATTACTGTAAAGGGTTCATCTT
ACTCTGGCATTCTGAGAATTAGACTGAAAGTTTAATTTCTGCAGTTCCTTCATATTAGATTCTTTCTTTGATGTTATA
ACACAAAGTCATTCTACTCAAATGTAATAAAATTGAGGCTCCACGGAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

FIGURE 32D

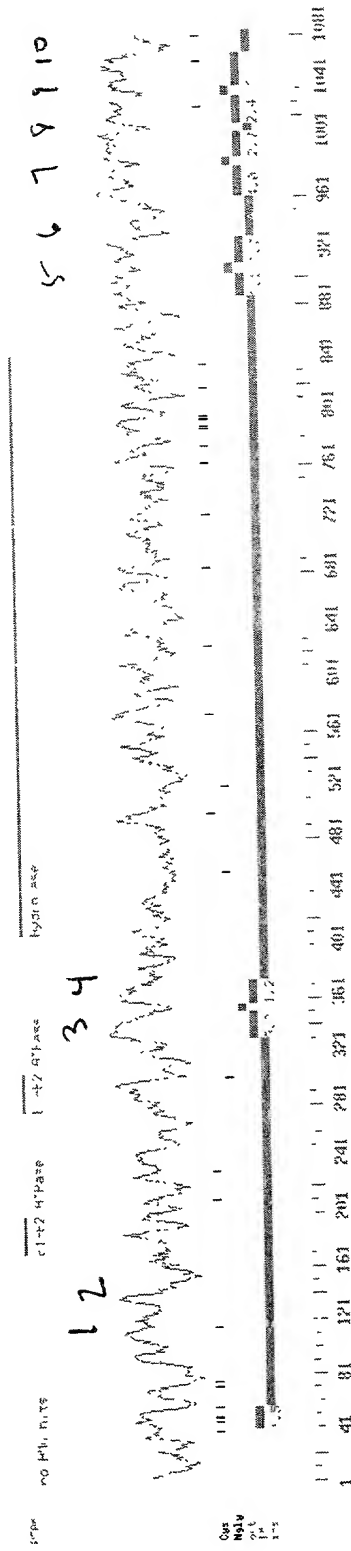


FIGURE 33

Protein Family / Domain Matches, HMMer version 2

Searching for complete domains in PFAM
 hmmpfam - search a single seq against HMM database
 HMMER 2.1.1 (Dec 1998)
 Copyright (C) 1992-1998 Washington University School of Medicine
 HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /prod/ddm/seqanal/PFAM/pfam6.4/Pfam
 Sequence file: /prod/ddm/wspace/orfanal/oa-script.16315.seq

Query: 67084FL

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
Hydrolase	haloacid dehalogenase-like hydrolase	19.2	0.0051	
1				
E1-E2_ATPase	E1-E2 ATPase	15.8	0.00087	
2				

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
E1-E2_ATPase	1/2	171	199	42	70	3.0	6.9
E1-E2_ATPase	2/2	277	305	105	133	13.0	0.0064
Hydrolase	1/1	410	843	1	184	19.2	0.0051

Alignments of top-scoring domains:

E1-E2_ATPase: domain 1 of 2, from 171 to 199: score 3.0, E = 6.9

```
*->keeeipaeLvpGDIVevkpGdrVPADgr<-*
      + +++++++GD+++v+  r+P D++
67084FL 171  GKVQVKSSDIQVGDLLIVEKNQRIPSDMV 199
```

E1-E2_ATPase: domain 2 of 2, from 277 to 305: score 13.0, E = 0.0064

```
*->lergnmVfaGTLvvsGsltgVtGddT<-*
      l + n+++a+T+v sG+ +gvV+ tG++T
67084FL 277  LSIENTLWASTIVASGTVIGVVIYTGKET 305
```

Hydrolase: domain 1 of 1, from 410 to 843: score 19.2, E = 0.0051

```
*->ikavvFDkDGTltdgkeppiaeaivealrelgl.....apleevekl
      + ++ Dk+GTLt+  + i + +  +g ++ +++ ++ ++
67084FL 410  LVYLLTDKTGTLTQ--NEMIFKRLHLGTVSYGAdtmdelQSHVRDSY 454

      lgrgl.g.....erilleggltaell.....
      ++++++++r+++++ + ++ +++ ++  +  ++ ++
67084FL 455  SQMQSqAggnntgstpLRKAQSSAPKVRKSvssriheavkaivlchnvtp 504

      .....
      +++ + +++++ + + + +++++ +++++ +++++ + +
67084FL 505  vyesragvteetefaeadqdfsdenrtyqasspdevalvqwtesvgltlv 554

      .....
      +++ ++ +++++ +  + +++++ + +++++ + +
67084FL 555  srdltsmqlktpsgqvlscilqlfpftseskrmgvivrdestaeitfym 604

      .....
      ++ +  ++  ++ +++ ++ +++ ++  ++ +++ + +++
67084FL 605  kgadvamspivqyndwleeecgnmareglrtlvakkalteeqyqdfesr 654

      .....ld.evlglial.dklypgarealkLk
      + + + +++++ +  ++ +++ e+l+l ++d+l ++r++l+ L+
67084FL 655  ytgaklsmhdrslkvaavvesleReMELLCLTGVeDQLQADVVRPTLEMLR 704
```

FIGURE 34A

```

erGikvailTngdr.naealle.....
+Gik+++lT++ ++a+ ++++++ +++++ + ++ +++++ + + +
67084FL 705 NAGIKIWMLTGDKLeTATCIAKsshlvsrtqdihi frqvt srgeahleln 754

.....algla.lfdaivdsdevggvgpvvvgKPkpe
+++++ +++++ +l + +++++ +++++ +VV+ + +p
67084FL 755 afrkrkhdcalvisgdslevCLK-YyEHEFVELACQCP---AVVCCRCSP 800

ifllalerlgvkpeevgpkvlmvGDginDapalaaAGvgvamngg<-*
+ +++ l+ + +++++GDg nD+ ++ aA++g+ +
67084FL 801 QKARIVTLLQQHTGRR---TCAIGDGGNDVSMIQAADCGIGIEGKE 843

```

FIGURE 34B

TCF 829200

LUSTAL W (1.74) multiple sequence alignment

bh67084FL AT2B	MPLMMSEEGFENEESDYHTLPRARIMQRKRGLEWVCDGWKFLCTSCCGWLINICRRKKE MPLMMSEEGFENDES DYHTLPRARITRRKRGLEWVCGGWKFLCTSCCDWLINVCQRKKE *****:*****:*****:*****:*****:*****: TM1
bh67084FL AT2B	LKARTVWLGCPKCEEKHPRNSIKNQKYNVFTFIPGVLYEQFKFFLNLYFLVISCQSOFVP LKARTVWLGCPKCEEKHPRNSIKNQKYNVFTFIPGVLYEQFKFFLNLYFLVVSCQSOFVP *****:*****:*****:*****:*****:*****: TM2
bh67084FL AT2B	ALKIGYLYTYWAPLGFVLAVTMTREAEIDEFRRFQORDKEVNSQLYSKLTVRGKVQVKSSDI ALKIGYLYTYWAPLGFVLAVTIAREAEIDEFRRFQORDKEMNSQLYSKLTVRGKVQVKSSDI *****:*****:*****:*****:*****:*****: Phospholipid Transport
bh67084FL AT2B	QVGDLIIVEKNQRIPSDMVFLRTSEKAGSCFIRIDQDGETDWKLKVAVSCTQQLPALGD QVGDLIIVEKNQRIPSDMVFLRTSEKAGSCFIRIDQDGETDWKLKVAVSCTQRLPALGD *****:*****:*****:*****:*****:*****:
Fbh67084FL AT2B	LFSISAYVYAQKPQMDIHSFEGTFTREDSDPPIHESLSIENTLWASTIVASGTVIGVVIY LFSISAYVYAQKPQLDIHSFEGTFTREDSDPPIHESLSIENTLWASTIVASGTVIGVVIY *****:*****:*****:*****:*****:*****: TM3
Fbh67084FL AT2B	TGKETRSVMNTSNPKNKVGLLDLELNRLTKALFLALVALSIVMVTLOGFVGPWYRNLFRF TGKETRSVMNTSNPNKVGLLDLELNQLTKALFLALVLSVVMVTLOGFAGPWYRNLFRF *****:*****:*****:*****:*****:*****: TM4
Fbh67084FL AT2B	LLLSFYIIPISLRVNLDMGKAVYGWMMMKDENIPGTVVRTSTIPEELGRLVYLLTDKTGT LLLSFYIIPISLRVNLDMGKAAYGWMIMKDENIPGTVVRTSTIPEELGRLVYLLTDKTGT *****:*****:*****:*****:*****:*****: Phospholipid Transport
Fbh67084FL AT2B	LTQONEMIFKRLHLGTVSYGADTMDEIQSHVRDSYSQMSQAGGNNTGSTPLRKAQSSAPK LTQONEMVFKRLHLGTVSYGTDTMDEIQSHVLNSYLQVHSQPSGHNPPSAPLRRSQSSTPK *****:*****:*****:*****:*****:*****: TM5
Fbh67084FL AT2B	VRKSVSSRIHEAVKAIVLCHNVTPVYESRAGVTEETFAEADQDFSDENRTYQASSPDEV VKKSVSSRIHEAVKAIALCHNVTPVYEARAGITGETFAEADQDFSDENRTYQASSPDEV *:*****:*****:*****:*****:*****:*****:
Fbh67084FL AT2B	ALVQWTESVGLTLVSRDLTSMQLKTPSGQVLSFCILQLFPFTSESKRMGIIVRDESTAEI ALVRWTESVGLTLVSRDLASMQLKTPSGQVLTTCILQMFPTSESKRMGIIVRDESTAEI ***:*****:*****:*****:*****:*****:
Fbh67084FL AT2B	TFYMKGADVAMSPIVQYNDWLEEECGNMAREGLRTLVAKKALTEEQYQDFESRYTQAKL TFYMKGADVAMSTIVQYNDWLEEECGNMAREGLRTLVAKRRLTEEQYQDFESRYSQAKL *****:*****:*****:*****:*****:*****:
Fbh67084FL AT2B	SMHDRSLKVAAVVESLEREMELLCLTGVEDQLQADVPTLEMLRNAGIKIWMLTGDKLET SIHDRALKVAAVVESLEREMELLCLTGVEDQLQADVPTLEMLRNAGIKIWMLTGDKLET *:***:*****:*****:*****:*****:*****:
Fbh67084FL AT2B	ATCIAKSSHLVSRTQDIHIFRQVTSRGEAHLELNAFRRKHDCALVISGDSLEVCLKYIEH ATCIAKSSHLVSRTQDIHVFRPVTSRGEAHLELNAFRRKHDCALVISGDSLEVCLRYIEH *****:*****:*****:*****:*****:*****: Phospholipid Transport
Fbh67084FL AT2B	EFVELACQCPAVVCCRCSPQKARIVTLLQOHTGRRTCAIGDGGNDVSMIQAADCGIGIE ELVELACQCPAVVCCRCSPTXKAHIVTLLRQHTKRRTCAIGDGGNDVSMIQAADCGIGIE *:*****:*****:*****:*****:*****:*****: TM6
Fbh67084FL	GKEGKQASLAADFSITQFRHIGRLLMVHGRNSYKRSALGQFVMHRGLIISTMQAVFSSV

FIGURE 35A

mAT2B

GKEGKQASLAADFSITQFRHIGRLLMVHGRNSYKRSAALGQFVMHRGLIISTMQAVFSSV

Tm6

Fbh67084FL

mAT2B

FYFASVPLYQGFLMVGYATIYTMFPVFSLVLDQDVKPEMAMLYPELYKDLTKGRSLSFKT
FYFASVPLYQGFLMVGYATIYTMFPVFSLVLDQDVKPEMAMLYPELYKDLTKGRSLSFKT

Tm7

Fbh67084FL

mAT2B

FLIWVLISIYQGGILMYGALVLFESFVHVVAISFTALILTELLMVALTVRTWHWLMVVA
FLIWVLISIYQGGILMYGALLFEDEFVHVVAISFTALILTELLXVALTIRTWHLMVVA

Tm9

Fbh67084FL

mAT2B

EFLSLGCVVSSLAFINEYF-----DVAFITTVTFLWKVSAITVVSCLPLYVIKY
EFLSLGCVVASLAFINEYFFGIGRVSFGAFLDVAFITTVTFLWKVSAITVVSCLPLYVIKY

Tm10

Fbh67084FL

mAT2B

LRRKSSPPSYCKLAS
LKRKLSPPSYSKLSS
: *:* *:*

FIGURE 35B

FIGURE 35B

Input file Fbh67084alt; Output File Fbh67084alt.tra
Sequence length 4231

GGAGTCGACCCACGCGTCCGCATTGAGACAATGCCTCCACAAATACTTGATGCAAAATTCAGTAAGACAGCACTTGTTG

AATCACCATTATAGTTTCTGACAAATTGTTCTCAAAAAGGTACCAGCTGGAGGATGAGTCTGCGCATTGATGAA

M	P	L	M	M	S	E	E	G	F	E	N	E	E	S	D	Y	H	T	L	20
ATG	CCA	CTA	ATG	ATG	TCT	GAA	GAA	GGC	TTT	GAG	AAT	GAG	GAA	AGT	GAT	TAC	CAC	ACC	TTA	60
P	R	A	R	I	M	Q	R	K	R	G	L	E	W	F	V	C	D	G	W	40
CCA	CGA	GCC	AGG	ATA	ATG	CAA	AGG	AAA	AGA	GGA	CTG	GAG	TGG	TTT	GTC	TGT	GAT	GGC	TGG	120
K	F	L	C	T	S	C	C	G	W	L	I	N	I	C	R	R	K	K	E	60
AAG	TTC	CTC	TGT	ACC	AGT	TGC	TGT	GGT	TGG	CTG	ATA	AAT	ATT	TGT	CGA	AGA	AAG	AAA	GAG	180
L	K	A	R	T	V	W	L	G	C	P	E	K	C	E	E	K	H	P	R	80
CTG	AAA	GCT	CGC	ACA	GTA	TGG	CTT	GGA	TGT	CCT	GAA	AAG	TGT	GAA	GAA	AAA	CAT	CCC	AGG	240
N	S	I	K	N	Q	K	Y	N	V	F	T	F	I	P	G	V	L	Y	E	100
AAT	TCT	ATA	AAA	AAT	CAA	AAA	TAC	AAT	GTG	TTT	ACC	TTT	ATA	CCT	GGG	GTT	TTG	TAT	GAA	300
Q	F	K	F	F	L	N	L	Y	F	L	V	I	S	C	S	Q	F	V	P	120
CAA	TTC	AAG	TTT	TTC	TTG	AAT	CTC	TAT	TTT	CTA	GTG	ATA	TCC	TGC	TCA	CAG	TTT	GTA	CCA	360
A	L	K	I	G	Y	L	Y	T	Y	W	A	P	L	G	F	V	L	A	V	140
GCA	TTG	AAA	ATA	GGC	TAT	CTC	TAC	ACC	TAC	TGG	GCT	CCT	CTG	GGA	TTT	GTC	TTG	GCT	GTT	420
T	M	T	R	E	A	I	D	E	F	R	R	F	Q	R	D	K	E	V	N	160
ACT	ATG	ACA	CGG	GAA	GCA	ATT	GAT	GAA	TTT	CGG	CGT	TTT	CAG	CGT	GAC	AAG	GAA	GTG	AAT	480
S	Q	L	Y	S	K	L	T	V	R	G	K	V	Q	V	K	S	S	D	I	180
TCA	CAA	CTA	TAT	AGC	AAG	CTT	ACA	GTA	AGA	GGT	AAA	GTG	CAA	GTT	AAG	AGT	TCA	GAC	ATA	540
Q	V	G	D	L	I	I	V	E	K	N	Q	R	I	P	S	D	M	V	F	200
CAA	GTT	GGA	GAC	CTC	ATC	ATA	GTG	GAA	AAG	AAT	CAA	AGA	ATT	CCA	TCG	GAC	ATG	GTG	TTT	600
L	R	T	S	E	K	A	G	S	C	F	I	R	T	D	Q	L	D	G	E	220
CTT	AGG	ACT	TCA	GAA	AAA	GCA	GGT	TCG	TGT	TTT	ATT	CGA	ACT	GAT	CAA	CTA	GAT	GGT	GAA	660
T	D	W	K	L	K	V	A	V	S	C	T	Q	Q	L	P	A	L	G	D	240
ACT	GAC	TGG	AAG	CTG	AAG	GTG	GCA	GTG	AGC	TGC	ACG	CAA	CAG	CTG	CCG	GCT	CTG	GGG	GAC	720
L	F	S	I	S	A	Y	V	Y	A	Q	K	P	Q	M	D	I	H	S	F	260
CTT	TTT	TCT	ATC	AGT	GCT	TAT	GTT	TAT	GCT	CAG	AAA	CCA	CAA	ATG	GAC	ATT	CAC	AGT	TTC	780
E	G	T	F	T	R	E	D	S	D	P	P	I	H	E	S	L	S	I	E	280
GAA	GGC	ACA	TTT	ACC	AGG	GAA	GAC	AGT	GAC	CCG	CCC	ATT	CAT	GAA	AGT	CTC	AGC	ATA	GAA	840
N	T	L	W	A	S	T	I	V	A	S	G	T	V	I	G	V	V	I	Y	300
AAT	ACA	TTG	TGG	GCA	AGC	ACC	ATT	GTT	GCA	TCA	GGT	ACT	GTA	ATA	GGT	GTT	GTC	ATT	TAT	900
T	G	K	E	T	R	S	V	M	N	T	S	N	P	K	N	K	V	G	L	320
ACC	GGA	AAA	GAG	ACT	CGA	AGT	GTA	ATG	AAC	ACA	TCC	AAT	CCA	AAA	AAT	AAG	GTT	GGT	TTG	960
L	D	L	E	L	N	R	L	T	K	A	L	F	L	A	L	V	A	L	S	340
TTG	GAC	CTT	GAA	CTC	AAT	CGG	CTG	ACG	AAA	GCG	CTA	TTT	TTG	GCT	TTA	GTT	GCT	CTT	TCC	1020

FIGURE 36A

I	V	M	V	T	L	Q	G	F	V	G	P	W	Y	R	N	L	F	R	F	360
ATT	GTT	ATG	GTA	ACC	TTA	CAA	GGA	TTT	GTG	GGT	CCA	TGG	TAC	CGC	AAT	CTT	TTT	CGG	TTC	1080
L	L	L	F	S	Y	I	I	P	I	S	L	R	V	N	L	D	M	G	K	380
CTT	CTC	CTC	TTT	TCT	TAC	ATC	ATT	CCC	ATA	AGT	TTG	CGT	GTG	AAC	TTG	GAC	ATG	GGC	AAA	1140
A	V	Y	G	W	M	M	M	K	D	E	N	I	P	G	T	V	V	R	T	400
GCG	GTG	TAT	GGA	TGG	ATG	ATG	ATG	AAA	GAT	GAG	AAC	ATC	CCT	GGC	ACG	GTC	GTT	CGG	ACC	1200
S	T	I	P	E	E	L	G	R	L	V	Y	L	L	T	D	K	T	G	T	420
AGC	ACT	ATC	CCA	GAG	GAA	CTT	GGG	CGC	CTG	GTG	TAT	TTA	TTG	ACA	GAC	AAA	ACA	GGA	ACC	1260
L	T	Q	N	E	M	I	F	K	R	L	H	L	G	T	V	S	Y	G	A	440
CTC	ACC	CAG	AAT	GAA	ATG	ATA	TTT	AAG	CGG	CTG	CAC	CTG	GGC	ACC	GTG	TCC	TAT	GGC	GCC	1320
D	T	M	D	E	I	Q	S	H	V	R	D	S	Y	S	Q	M	Q	S	Q	460
GAC	ACG	ATG	GAT	GAG	ATC	CAG	AGC	CAT	GTC	AGG	GAC	TCC	TAC	TCA	CAG	ATG	CAG	TCT	CAA	1380
A	G	G	N	N	T	G	S	T	P	L	R	K	A	Q	S	S	A	P	K	480
GCT	GGT	GGA	AAC	AAT	ACT	GGT	TCA	ACT	CCA	CTA	AGA	AAA	GCC	CAA	TCT	TCA	GCT	CCC	AAA	1440
V	R	K	S	V	S	S	R	I	H	E	A	V	K	A	I	V	L	C	H	500
GTT	AGG	AAA	AGT	GTC	AGT	AGT	CGA	ATC	CAT	GAA	GCC	GTG	AAA	GCC	ATC	GTG	CTG	TGT	CAC	1500
N	V	T	P	V	Y	E	S	R	A	G	V	T	E	E	T	E	F	A	E	520
AAC	GTG	ACC	CCC	GTG	TAT	GAG	TCT	CGG	GCC	GGC	GTT	ACT	GAG	GAG	ACT	GAG	TTC	GCA	GAG	1560
A	D	Q	D	F	S	D	E	N	R	T	Y	Q	A	S	S	P	D	E	V	540
GCT	GAC	CAA	GAC	TTC	AGT	GAT	GAG	AAT	CGC	ACC	TAC	CAG	GCT	TCC	AGC	CCG	GAT	GAG	GTC	1620
A	L	V	Q	W	T	E	S	V	G	L	T	L	V	S	R	D	L	T	S	560
GCT	CTG	GTG	CAG	TGG	ACA	GAG	AGT	GTG	GGC	CTC	ACG	CTG	GTC	AGC	AGG	GAC	CTC	ACC	TCC	1680
M	Q	L	K	T	P	S	G	Q	V	L	S	F	C	I	L	Q	L	F	P	580
ATG	CAG	CTG	AAG	ACC	CCC	AGT	GGC	CAG	GTC	CTC	AGC	TTC	TGC	ATT	CTG	CAG	CTG	TTT	CCC	1740
F	T	S	E	S	K	R	M	G	V	I	V	R	D	E	S	T	A	E	I	600
TTC	ACC	TCC	GAG	AGC	AAG	CGG	ATG	GGC	GTC	ATC	GTC	AGG	GAT	GAA	TCC	ACG	GCA	GAA	ATC	1800
T	F	Y	M	K	G	A	D	V	A	M	S	P	I	V	Q	Y	N	D	W	620
ACA	TTC	TAC	ATG	AAG	GGC	GCT	GAC	GTG	GCC	ATG	TCT	CCT	ATC	GTG	CAG	TAT	AAT	GAC	TGG	1860
L	E	E	E	C	G	N	M	A	R	E	G	L	R	T	L	V	V	A	K	640
CTG	GAA	GAG	GAG	TGC	GGA	AAC	ATG	GCT	CGC	GAA	GGA	CTG	CGG	ACC	CTC	GTG	GTT	GCA	AAG	1920
K	A	L	T	E	E	Q	Y	Q	D	F	E	S	R	Y	T	Q	A	K	L	660
AAG	GCG	TTG	ACA	GAG	GAG	CAG	TAC	CAG	GAC	TTT	GAG	AGC	CGA	TAC	ACT	CAA	GCC	AAG	CTG	1980
S	M	H	D	R	S	L	K	V	A	A	V	V	E	S	L	E	R	E	M	680
AGC	ATG	CAC	GAC	AGG	TCC	CTC	AAG	GTG	GCC	GCG	GTA	GTC	GAG	AGC	CTG	GAG	AGG	GAG	ATG	2040
E	L	L	C	L	T	G	V	E	D	Q	L	Q	A	D	V	R	P	T	L	700
GAA	CTG	CTG	TGC	CTC	ACC	GGC	GTG	GAG	GAC	CAG	CTG	CAG	GCA	GAC	GTG	CGG	CCC	ACG	CTG	2100
E	M	L	R	N	A	G	I	K	I	W	M	L	T	G	D	K	L	E	T	720
GAG	ATG	CTG	CGC	AAC	GCC	GGG	ATC	AAG	ATA	TGG	ATG	CTA	ACA	GGC	GAT	AAA	CTC	GAG	ACA	2160

FIGURE 36B

A	T	C	I	A	K	S	S	H	L	V	S	R	T	Q	D	I	H	I	F	740
GCT	ACC	TGC	ATT	GCC	AAA	AGT	TCA	CAT	CTC	GTG	TCT	AGA	ACA	CAA	GAT	ATT	CAT	ATT	TTC	2220
R	Q	V	T	S	R	G	E	A	H	L	E	L	N	A	F	R	R	K	H	760
AGA	CAG	GTA	ACC	AGT	CGG	GGA	GAG	GCA	CAT	TTG	GAG	CTG	AAT	GCA	TTT	CGA	AGG	AAG	CAT	2280
D	C	A	L	V	I	S	G	D	S	L	E	V	C	L	K	Y	Y	E	H	780
GAT	TGT	GCA	CTA	GTC	ATA	TCT	GGG	GAC	TCT	CTG	GAG	GTT	TGT	CTA	AAG	TAC	TAC	GAG	CAT	2340
E	F	V	E	L	A	C	Q	C	P	A	V	V	C	C	R	C	S	P	T	800
GAA	TTT	GTG	GAG	CTG	GCC	TGC	CAG	TGC	CCT	GCC	GTG	GTT	TGC	TGC	CGC	TGC	TCA	CCC	ACC	2400
Q	K	A	R	I	V	T	L	L	Q	Q	H	T	G	R	R	T	C	A	I	820
CAG	AAG	GCC	CGC	ATT	GTG	ACA	CTG	CTG	CAG	CAG	CAC	ACA	GGG	AGA	CGC	ACC	TGC	GCC	ATC	2460
G	D	G	G	N	D	V	S	M	I	Q	A	A	D	C	G	I	G	I	E	840
GGT	GAT	GGA	GGA	AAT	GAT	GTC	AGC	ATG	ATT	CAG	GCA	GCA	GAC	TGT	GGG	ATT	GGG	ATT	GAG	2520
G	K	E	G	K	Q	A	S	L	A	A	D	F	S	I	T	Q	F	R	H	860
GGA	AAG	GAG	GGT	AAA	CAG	GCC	TCG	CTG	GCG	GCC	GAC	TTC	TCC	ATC	ACG	CAG	TTC	CGG	CAC	2580
I	G	R	L	L	M	V	H	G	R	N	S	Y	K	R	S	A	A	L	G	880
ATA	GGC	AGG	CTG	CTC	ATG	GTG	CAC	GGG	CGG	AAC	AGC	TAC	AAG	AGG	TCG	GCG	GCA	CTC	GGC	2640
Q	F	V	M	H	R	G	L	I	I	S	T	M	Q	A	V	F	S	S	V	900
CAG	TTC	GTC	ATG	CAC	AGG	GGC	CTT	ATC	ATC	TCC	ACC	ATG	CAG	GCT	GTG	TTT	TCC	TCA	GTC	2700
F	Y	F	A	S	V	P	L	Y	Q	G	F	L	M	V	G	Y	A	T	I	920
TTC	TAC	TTC	GCA	TCC	GTC	CCT	TTG	TAT	CAG	GGC	TTC	CTC	ATG	GTG	GGG	TAT	GCC	ACC	ATA	2760
Y	T	M	F	P	V	F	S	L	V	L	D	Q	D	V	K	P	E	M	A	940
TAC	ACC	ATG	TTC	CCA	GTG	TTC	TCC	TTA	GTG	CTG	GAC	CAG	GAC	GTG	AAG	CCA	GAG	ATG	GCG	2820
M	L	Y	P	E	L	Y	K	D	L	T	K	G	R	S	L	S	F	K	T	960
ATG	CTC	TAC	CCG	GAG	CTG	TAC	AAG	GAC	CTC	ACC	AAG	GGA	AGA	TCC	TTG	TCC	TTC	AAA	ACC	2880
F	L	I	W	V	L	I	S	I	Y	Q	G	G	I	L	M	Y	G	A	L	980
TTC	CTC	ATC	TGG	GTT	TTA	ATA	AGT	ATT	TAC	CAA	GGC	GGC	ATC	CTC	ATG	TAT	GGG	GCC	CTG	2940
V	L	F	E	S	E	F	V	H	V	V	A	I	S	F	T	A	L	I	L	1000
GTG	CTC	TTC	GAG	TCT	GAG	TTC	GTC	CAC	GTG	GTG	GCC	ATC	TCC	TTC	ACC	GCA	CTG	ATC	CTG	3000
T	E	L	L	M	V	A	L	T	V	R	T	W	H	W	L	M	V	V	A	1020
ACC	GAG	CTG	CTG	ATG	GTG	GCG	CTG	ACC	GTC	CGC	ACG	TGG	CAC	TGG	CTG	ATG	GTG	GTG	GCC	3060
E	F	L	S	L	G	C	Y	V	S	S	L	A	F	L	N	E	Y	F	G	1040
GAG	TTC	CTC	AGC	TTA	GGC	TGC	TAC	GTG	TCC	TCA	CTC	GCT	TTT	CTC	AAT	GAA	TAT	TTT	GGT	3120
I	G	R	V	S	F	G	A	F	L	D	V	A	F	I	T	T	V	T	F	1060
ATA	GGC	AGA	GTG	TCT	TTT	GGA	GCT	TTC	TTA	GAT	GTT	GCC	TTT	ATC	ACC	ACC	GTG	ACC	TTC	3180
L	W	K	V	S	A	I	T	V	V	S	C	L	P	L	Y	V	L	K	Y	1080
CTG	TGG	AAA	GTG	TCG	GCG	ATC	ACC	GTG	GTC	AGC	TGC	CTC	CCG	CTG	TAT	GTC	CTC	AAG	TAC	3240
L	R	R	K	S	S	P	P	S	Y	C	K	L	A	S	*					1096
CTG	AGG	CGC	AAG	TCT	TCT	CCT	CCC	AGC	TAC	TGC	AAG	CTG	GCC	TCC	TAA					3288

GGGGCTGTGCACCCCCAGCGGGCTGGCCCCAGCACCTTCTGCCCTTCCCAGCACCTTGTGCCCTTGCCAGTGAACGCAG

FIGURE 36C

1002463-121704

GGTTTGCCATTGCTACCAAGCAAGCACCAAGAAAGGGAGGGTACGCCAGGCGAGCCCAGGGCACAGATGCTGAGACA
GCCTCTCCTTCTCAGTGCAGGGACGTCACCCCTGCCAGGCAAGCCCAGGGCACAGATGCCAGGATGGCTTCTCCCTCTC
AGTGCAGGGCTTCACCCCTGCCAGGCAAGCCCAGGGCATAGATGCTGAGACAGCCTCTCCCTCTCAGTGCAGGGACGTC
ACCCCTGCCAGGCAAGCCCAGGGCACAGAGGCCGGGACGGCCTCTCCCTCTCAGTGTGAGGCTTCACCCATGCTAGGCA
AGCCCAGGGCACAGATGCCGGGATGGCCCTCCCTCTCAGTGCAGGGAACGTCACCCCTGCCAGGCAAGCCCAGGGCACACA
GATGCTGCGATGGCCTCTTCCTCTTAAGTGTGGGGCCTCACCCCTGCTTTTCTTTCTTTTTTGTATGTCAAATTGT
ATTTCCATATTGAAGCAGCTTGAGTTTCTACTGAAAATGAGCCCGAATTATTTCACTATTACTGTAAAGGGTTCATCTT
ACTCTGGCATTCTGAGAATTAGACTGAAAGTTAATTTCTGCAGTTCCTCATATTCAGATTCTTTCTTTGATGTTATA
ACACAAAGTCATTCTACTCAAATGTAATAAAATTGAGGCTCCACGGAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

FIGURE 36D

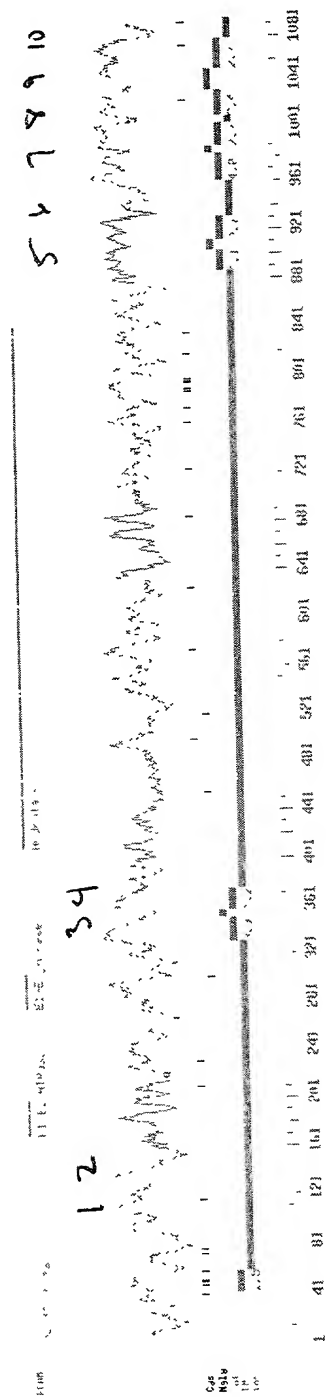


FIGURE 37

Protein Family / Domain Matches, HMMer version 2

Searching for complete domains in PFAM

hmmpfam - search a single seq against HMM database

HMMER 2.1.1 (Dec 1998)

Copyright (C) 1992-1998 Washington University School of Medicine

HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /prod/ddm/seqanal/PFAM/pfam6.4/Pfam
Sequence file: /prod/ddm/wspace/orfanal/oa-script.17118.seq

Query: 67084alt

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
Hydrolase	haloacid dehalogenase-like hydrolase	19.2	0.0051	1
E1-E2_ATPase	E1-E2 ATPase	15.8	0.00087	2

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
E1-E2_ATPase	1/2	171	199	..	42 70 ..	3.0	6.9
E1-E2_ATPase	2/2	277	305	..	105 133 ..	13.0	0.0064
Hydrolase	1/1	410	843	..	1 184 []	19.2	0.0051

Alignments of top-scoring domains:

E1-E2_ATPase: domain 1 of 2, from 171 to 199: score 3.0, E = 6.9

```
*->keeeipaeLvpGDIVevkpGdrVPADgr<-*
+ +++++++GD+++v+ r+P D++
67084alt 171 GKVQVKSSDIQVGDLIIVEKNQRIPSDMV 199
```

E1-E2_ATPase: domain 2 of 2, from 277 to 305: score 13.0, E = 0.0064

```
*->lerngmVfaGTLvvsGsiltgvVtatGddT<-*
l + n+++a+T+v sG+ +gvV+ tG++T
67084alt 277 LSIENTLWASTIVASGTVIGVVIYTGKET 305
```

Hydrolase: domain 1 of 1, from 410 to 843: score 19.2, E = 0.0051

```
*->ikavvFDkDGTLTdgtkeppiaeaivealrelgl.....apleevekl
+ ++ Dk+GTLt+ + i + + +g ++ ++ ++ ++
67084alt 410 LVYLLTDKTGTLTQ--NEMIFKRLHLGTVSYGAdtmdeIQSHVRDSY 454

lgrgl.g.....erillegltaell.....
+++++++R+++++ + ++ ++ ++ + ++ ++
67084alt 455 SQMQSqAggnntgstplRKAQSSAPKVRKsvsriheavkaivlchnvtp 504

.....
+++ + +++++ + + + +++++ +++++ +++++ + +
67084alt 505 vyesragvteetefaeadqdfsdnrttyqasspdevalvqwtesvgltlv 554

.....
+++ ++ +++++ + + +++++ + +++++ + +
67084alt 555 srdltsmqlktpsgqvlscilqlfpftseskrmgvivvrdetaeitfym 604

.....
++ + ++ ++ ++ ++ ++ ++ ++ ++ ++
67084alt 605 kgadvamspivqyndwleeeecnmareglrtlvakkalteeqyqdfesr 654

.....ld.evlglial.dklypgarealkK
+ + + +++ + ++ ++ e+l+l ++d+l ++r++l+ L+
67084alt 655 ytgaklsmhdsrikvaavvesleREmELLCLTGVeDQLQADVVRPTLEMLR 704
```

FIGURE 38A


```

erGikvailTngdr.naealle.....
+Gik++lT++ ++a+ ++++++ +++++ + ++ +++++ + + +
67084alt 705 NAGIKIWMLTGDKLeTATCIAKsshlvstqdihi frqvt srgeahleln 754

.....algla.lfdaivdsdevggv gpvvvgKPkpe
+++++ +++++ + +l + +++++ +++++ +vv+ + +p
67084alt 755 afrrkhdcalvisgdslevCLK-YyEHEFVELACQCP---AVVCCRCSP 800

ifllalerlgvkpeevgpkv lmvGDginDapalaaAGvgvamngg<-*
+ +++ l+ + +++++GDg nD+ ++ aA++g+ +
67084alt 801 QKARIVTLLQQHTGRR---TCAIGDGGNDVSMIQAADCGIGIEGKE 843

```

Top of page

FIGURE 38B

FIGURE 39A

mAT2B

GKEGKQASLAADFSITQFRHIGRLLMVHGRNSYKRSAALGQFVMHRGLIISTMQAVFSSV

TM6

Fbh67084alt
mAT2B

FYFASVPLYQCFLMVGYATIYTMFPVFSLVLDQDVKPEMAMLYPELYKDLTKGRSLSFKT
FYFASVPLYQCFLMVGYATIYTMFPVFSLVLDQDVKPEMAMLYPELYKDLTKGRSLSFKT

Fbh67084alt
mAT2B

FLIWVLISIYQGGILMYGALVLESEFVHVVAISFTALILTELLMVALTVRTWHLMVVA
FLIWVLISIYQGGILMYGALLFEDEFVHVVAISFTALILTELLXVALTIRTWHLMVVA

Fbh67084alt
mAT2B

EFLSLGCVSSLAFLNEYFGIGRVSFGAFLDVAFITTVTFLWKVSAITVVSCLPLYVIKY
EFLSLGCVVASLAFLNEYFGIGRVSFGAFLDVAFITTVTFLWKVSAITVVSCLPLYVIKY

Fbh67084alt
mAT2B

LRRKSSPPSYCKLAS
LKRKLSPPSYSKLSS
: * * * * : * *

TM7

TM8

TM9

TM10

FIGURE 39B